

APPLICATION
FOR
UNITED STATES LETTERS PATENT

TITLE: MOLECULAR SEQUENCE OF SWINE RETROVIRUS AND
METHODS OF USE

APPLICANT: JAY A. FISHMAN

CERTIFICATE OF MAILING BY EXPRESS MAIL

Express Mail Label No. EL 983 022 743 US

November 26, 2003
Date of Deposit

MOLECULAR SEQUENCE OF SWINE RETROVIRUS AND METHODS OF USE

This application is a divisional of U.S.S.N. 09/661,858, filed on September 14, 2000, which is a divisional of U.S.S.N. 08/766,528, filed on December 13, 1996, which is a continuation-in-part of U.S.S.N. 08/572,645, filed on December 14, 1995, the entire contents of which are hereby incorporated by reference.

Field of the Invention

The invention relates to porcine retroviral sequences, peptides encoded by porcine retroviral sequences, and methods of using the porcine retroviral nucleic acids and peptides.

Background of the Invention

Advances in solid organ transplantation and a chronic shortage of suitable organ donors have made xenotransplantation an attractive alternative to the use of human allografts. However, the potential for introduction of a new group of infectious diseases from donor animals into the human population is a concern with the use of these methods.

The term applied to the natural acquisition by humans of infectious agents carried by other species is zoonosis. The transplantation of infection from nonhuman species into humans is best termed "direct zoonosis" or "xenosis."

Nonhuman primates and swine have been considered the main potential sources of organs for xenotransplantation (Niekrasz et al. (1992) *Transplant Proc* 24:625; Starzl et al. (1993) *Lancet* 341:65; Murphy et al. (1970) *Trans Proc* 4:546; Brede and Murphy (1972) *Primates Med* 7:18; Cooper et al. In Xenotransplantation: The Transplantation of Organs and Tissues between Species, eds. Cooper et al. (1991) p. 457; R Y Calne (1970) *Transplant Proc* 2:550; H. Auchincloss, Jr. (1988) *Transplantation* 46:1; and Chiche et al. (1993) *Transplantation* 6:1418). The infectious disease issues for primates and swine are similar to those of human donors. The prevention of infection depends on the ability to predict, to recognize, and to prevent common infections in the immunocompromised transplantation recipient (Rubin et al. (1993) *Antimicrob Agents Chemother* 37:619). Because of the potential carriage by nonhuman primates of pathogens easily adopted to humans, ethical concerns, and the cost of maintaining large colonies of primates, other species have received consideration as organ donors (Brede and Murphy (1972) *Primates Med* 7:18; Van Der Riet et al. (1987) *Transplant Proc* 19:4069; Katler In Xenotransplantation: The Transplantation of Organs and Tissues between Species, eds. Cooper et al. (1991) p. 457; Metzger et al. (1981) *J Immunol* 127:769; McClure et al. (1987) *Nature* 330:487; Letvin et al. (1987) *J Infect Dis* 156:406; Castro et al. (1991) *Virology* 184:219; Benveniste and Todaro (1973) *Proc Natl Acad Sci USA* 70:3316; and Teich, in RNA Tumor viruses, eds. Weiss et. al. (1985) p. 25). The economic importance of swine and experience in studies of transplantation in the miniature swine model have allowed some of the potential pathogens associated with these animals to be defined (Niekrasz et al. (1992) *Transplant Proc* 24:625; Cooper et al. In Xenotransplantation: The Transplantation of Organs and Tissues between Species, eds.

Cooper et al. (1991) p. 457; and Leman et al. (1992) Diseases of Swine, 7th ed. Ames, Iowa:Iowa State University). Miniature swine have received consideration as organ donors because of a number of features of the species. The structure and function of the main pig organs are comparable to those of man. Swine attain body weights and organ sizes adequate to the provision of organs for human use. Lastly, veterinarians and commercial breeders have developed approaches to creation of specific-pathogen-free (SPF) swine with the ability to eliminate known pathogens from breeding colonies (Alexander et al. (1980) *Proc 6th Int Congr Pig Vet Soc*, Copenhagen; Betts (1961) *Vet Rec* 73:1349; Betts et al. (1960) *Vet Rec* 72:461; Caldwell et al. (1959) *J Am Vet Med Assoc* 135:504; and Yong (1964) *Adv Vet Sci* 9:61).

Concern exists over the transfer of porcine retroviruses by xenotransplantation (Smith (1993) *N Engl J Med* 328:141). Many of the unique properties of the retroviruses are due to the synthesis of a complementary DNA copy from the RNA template (by reverse transcriptase), and integration of this DNA into the host genome. The integrated retroviral copy (which is referred to as an endogenous copy or "provirus") can be transmitted via the germ line.

Summary of the Invention

In general, the invention features a purified swine or miniature swine retroviral nucleic acid, e.g., a Tsukuba nucleic acid, a purified miniature swine retroviral nucleic acid sequence of SEQ ID NO:1 or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement, and methods of their use in detecting the presence of porcine, e.g., miniature swine, retroviral sequences.

In another aspect, the invention features a purified nucleic acid, e.g., a probe or primer, which can specifically hybridize with a purified swine or miniature swine retroviral genome, e.g., a Tsukuba genome, the sequence of SEQ ID NO:1 or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement.

In preferred embodiments the nucleic acid is other than the entire retroviral genome of SEQ ID NO:1 or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement, e.g., it is at least 1 nucleotide longer, or at least 1 nucleotide shorter, or differs in sequence at at least one position, e.g., the nucleic acid is a fragment of the sequence of SEQ ID NO:1 or its complement SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement, or it includes sequence additional to that of SEQ ID NO:1, or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement.

In preferred embodiments, the nucleic acid has at least 60%, 70%, 72%, more preferably at least 85%, more preferably at least 90%, more preferably at least 95%, most preferably at least 98%, 99% or 100% sequence identity or homology with a sequence from SEQ ID NO:1 or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement.

In other embodiments: the sequence of the nucleic acid differs from the corresponding sequence of SEQ ID NO: 1 or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement, by 1, 2, 3, 4, or 5 base pairs; the sequence of the nucleic acid differs from the corresponding sequence of SEQ ID NO: 1 or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement, by at least 1, 2, 3, 4, or 5 base pairs but less than 6, 7, 8, 9, or 10 base pairs.

In other preferred embodiments: the nucleic acid is at least 10, more preferably at least 15, more preferably at least 20, most preferably at least 25, 30, 50, 100, 1000, 2000, 4000, 6000, or 8060 nucleotides in length; the nucleic acid is less than 15, more preferably less than 20, most preferably less than 25, 30, 50, 100, 1000, 2000, 4000, 6000, or 8060 nucleotides in length.

In yet other preferred embodiments: the nucleic acid can specifically hybridize with a translatable region of a miniature swine retroviral genome, e.g., the retroviral genome of SEQ ID NO: 1, or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement, e.g., a region from the gag, pol, or env gene; the probe or primer can specifically hybridize with an untranslated region of a miniature swine retroviral genome, e.g., the retroviral genome of SEQ ID NO: 1, or its complement SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement; the probe or primer can specifically hybridize with a non-conserved region of a miniature swine retroviral genome, e.g., the retroviral genome of SEQ ID NO: 1, or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement; the probe or primer can specifically hybridize with the highly conserved regions of a miniature swine retroviral genome, e.g., the retroviral genome of SEQ ID NO: 1, or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement.

In preferred embodiments, the primer is selected from the group consisting of SEQ ID NOs:4-74.

In preferred embodiments, hybridization of the probe to retroviral sequences can be detected by standard methods, e.g., by radiolabeled probes or by probes bearing nonradioactive markers such as enzymes or antibody binding sites. For example, a probe can be conjugated with an enzyme such as horseradish peroxidase, where the enzymatic activity of the conjugated enzyme is used as a signal for hybridization. Alternatively, the probe can be coupled to an epitope recognized by an antibody, e.g., an antibody conjugated to an enzyme or another marker.

In another aspect, the invention features a reaction mixture which includes a target nucleic acid, e.g., a human, swine, or a miniature swine nucleic acid, and a purified second nucleic acid, e.g., a probe or primer, as, e.g., is described herein, which specifically hybridizes with the sequence of SEQ ID NO:1 or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement, a swine or a miniature swine retroviral nucleic acid, e.g., a Tsukuba nucleic acid.

In preferred embodiments, the target nucleic acid: includes RNA; or includes DNA.

In preferred embodiments, the target nucleic acid includes: genomic DNA isolated from a miniature swine; RNA or cDNA, e.g., cDNA made from an RNA template, isolated from a miniature swine; DNA, RNA or cDNA, e.g., cDNA made from an RNA template, isolated from a miniature swine organ, e.g., a kidney; RNA, DNA or cDNA, e.g., cDNA made from an RNA template, isolated from a miniature swine potential donor organ; RNA, DNA or cDNA, e.g., cDNA made from an RNA template, isolated from a miniature swine organ which has been transplanted into a organ recipient, e.g., a xenogeneic recipient, e.g., a primate, e.g., a human.

10 In preferred embodiments, the target nucleic acid includes: genomic DNA isolated from a swine; RNA or cDNA, e.g., cDNA made from an RNA template, isolated from a swine; DNA, RNA or cDNA, e.g., cDNA made from an RNA template, isolated from a swine organ, e.g., a kidney; RNA, DNA or cDNA, e.g., cDNA made from an RNA template, isolated from a swine potential donor organ; RNA, DNA or cDNA, e.g., cDNA made from an RNA template, isolated from a swine organ which has been transplanted into a organ recipient, e.g., a xenogeneic recipient, e.g., a primate, e.g., a human.

15 In a preferred embodiment: the second nucleic acid is a porcine retroviral sequence, probe or primer, e.g., as described herein, e.g., a Tsukuba-1 retroviral sequence: the second nucleic acid is a sequence of SEQ ID NO:1 or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement, or a fragment of the sequence or complement at least 10, 20, or 30, basepairs in length.

20 In preferred embodiments, the second nucleic acid has at least 60%, 70%, 72%, more preferably at least 85%, more preferably at least 90%, more preferably at least 95%, most preferably at least 98%, 99% or 100% sequence identity or homology with a sequence from SEQ ID NO:1 or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement.

25 In other preferred embodiments: the second nucleic acid is at least 10, more preferably at least 15, more preferably at least 20, most preferably at least 25, 30, 50, 100, 1000, 2000, 4000, 6000, or 8060 nucleotides in length; the nucleic acid is less than 15, more preferably less than 20, most preferably less than 25, 30, 50, 100, 1000, 2000, 4000, 6000, or 8060 nucleotides in length; the second nucleic acid is a full length retroviral genome.

30 In preferred embodiments the second nucleic acid is: a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence which encodes a gag protein; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence from nucleotides 2452-4839 (e.g, from nucleotides 3112-4683) of SEQ ID NO:1, nucleotides 598-2169 (e.g, from nucleotides 598-2169) of SEQ ID NO:2, or nucleotides 585-2156 (e.g, from nucleotides 585-2156) of SEQ ID NO:3, or naturally occurring mutants thereof; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence which encodes a pol protein; a nucleic acid of at least 10 consecutive nucleotides of sense or

antisense sequence from nucleotides 4871-8060 of SEQ ID NO:1, nucleotides 2320-4737 of SEQ ID NO:2, or nucleotides 2307-5741 of SEQ ID NO:3, or naturally occurring mutants thereof; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence which encodes a env protein; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence from nucleotides 2-1999 (e.g, from nucleotides 86-1999) of SEQ ID NO:1, nucleotides 4738-6722 (e.g, from nucleotides 4738-6722) of SEQ ID NO:2, or nucleotides 5620-7533 of SEQ ID NO:3, or naturally occurring mutants thereof.

In another aspect, the invention features a method for screening a cell or a tissue, e.g., a cellular or tissue transplant, e.g., a xenograft, for the presence or expression of a swine or a miniature swine retrovirus or retroviral sequence, e.g., an endogenous miniature swine retrovirus. The method includes:

contacting a target nucleic acid from the tissue with a second sequence chosen from the group of: a sequence which can specifically hybridize to a porcine retroviral sequence; a sequence which can specifically hybridize to the sequence of SEQ ID NO:1 or its complement; a sequence which can specifically hybridize to the sequence of SEQ ID NO:2 or its complement; a sequence which can specifically hybridize to the sequence of SEQ ID NO:3 or its complement; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence which encodes a gag protein; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence from nucleotides 2452-4839 (e.g, from nucleotides 3112-4683) of SEQ ID NO:1, nucleotides 598-2169 (e.g, from nucleotides 598-2169) of SEQ ID NO:2, or nucleotides 585-2156 (e.g, from nucleotides 585-2156) of SEQ ID NO:3, or naturally occurring mutants thereof; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence which encodes a pol protein; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence from nucleotides 4871-8060 of SEQ ID NO:1, nucleotides 2320-4737 of SEQ ID NO:2, or nucleotides 2307-5741 of SEQ ID NO:3, or naturally occurring mutants thereof; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence which encodes a env protein; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence from nucleotides 2-1999 (e.g, from nucleotides 86-1999) of SEQ ID NO:1, nucleotides 4738-6722 (e.g, from nucleotides 4738-6722) of SEQ ID NO:2, or nucleotides 5620-7533 of SEQ ID NO:3, or naturally occurring mutants thereof; a swine or miniature swine retroviral nucleic acid; or a Tsukuba nucleic acid under conditions in which hybridization can occur, hybridization being indicative of the presence or expression of an endogenous miniature swine retrovirus or retroviral sequence in the tissue or an endogenous swine retrovirus in the tissue.

In preferred embodiments, the method further includes amplifying the target nucleic acid with primers which specifically hybridize to the sequence of SEQ ID NO:1 or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement.

In preferred embodiments, the tissue or cellular transplant is selected from the group consisting of: heart, lung, liver, bone marrow, kidney, brain cells, neural tissue, pancreas or pancreatic cells, thymus, or intestinal tissue.

In other preferred embodiments, the target nucleic acid is: DNA; RNA; or cDNA.

5 In other preferred embodiments, the target nucleic acid is taken from: a tissue sample, or a blood sample, e.g., a tissue biopsy sample, e.g., a tissue sample suitable for *in situ* hybridization or immunohistochemistry.

In preferred embodiments, the target nucleic acid includes: genomic DNA isolated from a miniature swine; RNA or cDNA, e.g., cDNA made from an RNA template, isolated from a miniature swine; DNA, RNA or cDNA, e.g., cDNA made from an RNA template, isolated from a miniature swine organ, e.g., a kidney; RNA, DNA or cDNA, e.g., cDNA made from an RNA template, isolated from a miniature swine potential donor organ; RNA, DNA or cDNA, e.g., cDNA made from an RNA template, isolated from a miniature swine organ which has been transplanted into a organ recipient, e.g., a xenogeneic recipient, e.g., a primate, e.g., a human.

15 In preferred embodiments, the target nucleic acid includes: genomic DNA isolated from a swine; RNA or cDNA, e.g., cDNA made from an RNA template, isolated from a swine; DNA, RNA or cDNA, e.g., cDNA made from an RNA template, isolated from a swine organ, e.g., a kidney; RNA, DNA or cDNA, e.g., cDNA made from an RNA template, isolated from a swine potential donor organ; RNA, DNA or cDNA, e.g., cDNA made from an RNA template, isolated from a swine organ which has been transplanted into a organ recipient, e.g., a recipient swine or a xenogeneic recipient, e.g., a primate, e.g., a human.

20 In a preferred embodiment the target nucleic acid is RNA, or a nucleic acid amplified from RNA in the tissue, and hybridization is correlated with expression of an endogenous miniature swine retrovirus or retroviral sequence or an endogenous swine retrovirus.

In a preferred embodiment the target nucleic acid is DNA, or a nucleic acid amplified from DNA in the tissue, and hybridization is correlated with the presence of an endogenous miniature swine retrovirus or an endogenous swine retrovirus.

30 In preferred embodiments, the second nucleic acid has at least 60%, 70%, 72%, more preferably at least 85%, more preferably at least 90%, more preferably at least 95%, most preferably at least 98%, 99% or 100% sequence identity or homology with a sequence from SEQ ID NO:1 or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement.

35 In other preferred embodiments: the second nucleic acid is at least 10, more preferably at least 15, more preferably at least 20, most preferably at least 25, 30, 50, 100, 1000, 2000, 4000, 6000, or 8060 nucleotides in length; the nucleic acid is less than 15, more preferably less than 20, most preferably less than 25, 30, 50, 100, 1000, 2000, 4000, 6000, or 8060 nucleotides in length; the second nucleic acid is a full length retroviral genome.

In another aspect, the invention features a method of screening a porcine derived cell or tissue for the presence of an activatable porcine retrovirus, e.g., an activatable porcine provirus. The method includes:

stimulating a porcine derived cell or tissue with a treatment which can activate a
5 retrovirus;

contacting a target nucleic acid from the porcine derived cell or tissue with a second sequence chosen from the group of: a sequence which can specifically hybridize to a porcine retroviral sequence; a sequence which can specifically hybridize to the sequence of SEQ ID NO:1 or its complement; a sequence which can specifically hybridize to the sequence of SEQ
10 ID NO:2 or its complement; a sequence which can specifically hybridize to the sequence of SEQ ID NO:3 or its complement; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence which encodes a gag protein; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence from nucleotides 2452-4839 (e.g, from nucleotides 3112-4683) of SEQ ID NO:1, nucleotides 598-2169 (e.g, from nucleotides
15 598-2169) of SEQ ID NO:2, or nucleotides 585-2156 (e.g, from nucleotides 585-2156) of SEQ ID NO:3, or naturally occurring mutants thereof; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence which encodes a pol protein; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence from nucleotides 4871-8060 of SEQ ID NO:1, nucleotides 2320-4737 of SEQ ID NO:2, or
20 nucleotides 2307-5741 of SEQ ID NO:3, or naturally occurring mutants thereof; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence which encodes a env protein; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence from nucleotides 2-1999 (e.g, from nucleotides 86-1999) of SEQ ID NO:1, nucleotides 4738-6722 (e.g, from nucleotides 4738-6722) of SEQ ID NO:2, or nucleotides
25 5620-7533 of SEQ ID NO:3, or naturally occurring mutants thereof; a swine or miniature swine retroviral nucleic acid; or a Tsukuba nucleic acid hybridization being indicative of the presence of an activatable porcine provirus in the porcine derived cell or tissue.

In preferred embodiments the treatment is: contact with a drug, e.g., a steroid or a cytotoxic agent, infection or contact with a virus, the induction of stress, e.g., nutritional
30 stress or immunologic stress, e.g., contact with a T-cell, e.g., a reactive T-cell.

In preferred embodiments, the method further includes amplifying the target nucleic acid with primers which specifically hybridize to the sequence of SEQ ID NO:1 or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement.

In other preferred embodiments, the target nucleic acid is taken from: a tissue sample,
35 or a blood sample, e.g., a tissue biopsy sample, e.g., a tissue sample suitable for *in situ* hybridization or immunohistochemistry.

In preferred embodiments, the target nucleic acid includes: genomic DNA isolated from a miniature swine; RNA or cDNA, e.g., cDNA made from an RNA template, isolated from a miniature swine; DNA, RNA or cDNA, e.g., cDNA made from an RNA template,

isolated from a miniature swine organ, e.g., a kidney; RNA, DNA or cDNA, e.g., cDNA made from an RNA template, isolated from a miniature swine potential donor organ; RNA, DNA or cDNA, e.g., cDNA made from an RNA template, isolated from a miniature swine organ which has been transplanted into a organ recipient, e.g., a xenogeneic recipient, e.g., a primate, e.g., a human.

In preferred embodiments, the target nucleic acid includes: genomic DNA isolated from a swine; RNA or cDNA, e.g., cDNA made from an RNA template, isolated from a swine; DNA, RNA or cDNA, e.g., cDNA made from an RNA template, isolated from a swine organ, e.g., a kidney; RNA, DNA or cDNA, e.g., cDNA made from an RNA template, isolated from a swine potential donor organ; RNA, DNA or cDNA, e.g., cDNA made from an RNA template, isolated from a swine organ which has been transplanted into a organ recipient, e.g., a recipient swine or a xenogeneic recipient, e.g., a primate, e.g., a human.

In preferred embodiments, the second nucleic acid has at least 60%, 70%, 72%, more preferably at least 85%, more preferably at least 90%, more preferably at least 95%, most preferably at least 98%, 99% or 100% sequence identity or homology with a sequence from SEQ ID NO:1 or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement.

In other preferred embodiments: the second nucleic acid is at least 10, more preferably at least 15, more preferably at least 20, most preferably at least 25, 30, 50, 100, 1000, 2000, 4000, 6000, or 8060 nucleotides in length; the nucleic acid is less than 15, more preferably less than 20, most preferably less than 25, 30, 50, 100, 1000, 2000, 4000, 6000, or 8060 nucleotides in length; the second nucleic acid is a full length retroviral genome.

In another aspect, the invention features a method for screening a miniature swine genome or a swine genome for the presence of a porcine retrovirus or retroviral sequence, e.g., an endogenous porcine retrovirus. The method includes:

contacting the miniature swine (or swine) genomic DNA with a second sequence chosen from the group of: a sequence which can specifically hybridize to a porcine retroviral sequence; a sequence which can specifically hybridize to the sequence of SEQ ID NO:1 or its complement; a sequence which can specifically hybridize to the sequence of SEQ ID NO:2 or its complement; a sequence which can specifically hybridize to the sequence of SEQ ID NO:3 or its complement; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence which encodes a gag protein; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence from nucleotides 2452-4839 (e.g, from nucleotides 3112-4683) of SEQ ID NO:1, nucleotides 598-2169 (e.g, from nucleotides 598-2169) of SEQ ID NO:2, or nucleotides 585-2156 (e.g, from nucleotides 585-2156) of SEQ ID NO:3, or naturally occurring mutants thereof; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence which encodes a pol protein; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence from nucleotides 4871-8060 of SEQ ID NO:1, nucleotides 2320-4737 of SEQ ID NO:2, or nucleotides 2307-5741 of

SEQ ID NO:3, or naturally occurring mutants thereof; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence which encodes a env protein; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence from nucleotides 2-1999 (e.g, from nucleotides 86-1999) of SEQ ID NO:1, nucleotides 4738-6722 (e.g, from nucleotides 4738-6722) of SEQ ID NO:2, or nucleotides 5620-7533 of SEQ ID NO:3, or naturally occurring mutants thereof; a swine or miniature swine retroviral nucleic acid; or a Tsukuba nucleic acid under conditions in which the sequences can hybridize, hybridization being indicative of the presence of the endogenous porcine retrovirus or retroviral sequence in the miniature swine (or swine) genome.

10 In preferred embodiments, the method further includes amplifying all or a portion of the miniature swine (or swine) genome with primers which specifically hybridize to the sequence of SEQ ID NO:1 or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement.

15 In a preferred embodiment: the second nucleic acid is a porcine retroviral sequence, probe or primer, e.g., as described herein, e.g., a Tsukuba-1 retroviral sequence; the second nucleic acid is a sequence of SEQ ID NO:1 or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement, or a fragment of the sequence or complement at least 10, 20, or 30, basepairs in length.

20 In preferred embodiments, the second nucleic acid has at least 60%, 70%, 72%, more preferably at least 85%, more preferably at least 90%, more preferably at least 95%, most preferably at least 98%, 99% or 100% sequence identity or homology with a sequence from SEQ ID NO:1 or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement.

25 In other preferred embodiments: the second nucleic acid is at least 10, more preferably at least 15, more preferably at least 20, most preferably at least 25, 30, 50, 100, 1000, 2000, 4000, 6000, or 8060 nucleotides in length; the nucleic acid is less than 15, more preferably less than 20, most preferably less than 25, 30, 50, 100, 1000, 2000, 4000, 6000, or 8060 nucleotides in length; the second nucleic acid is a full length retroviral genome.

30 In another aspect, the invention features a method for screening a genetically modified miniature swine or a genetically modified swine for the presence or expression of a miniature swine or swine retrovirus or retroviral sequence, e.g., an endogenous miniature swine retrovirus. The method includes:

35 contacting a target nucleic acid from the genetically modified miniature swine or swine with a second sequence chosen from the group of: a sequence which can specifically hybridize to a porcine retroviral sequence; a sequence which can specifically hybridize to the sequence of SEQ ID NO:1 or its complement; a sequence which can specifically hybridize to the sequence of SEQ ID NO:2 or its complement; a sequence which can specifically hybridize to the sequence of SEQ ID NO:3 or its complement; a nucleic acid of at least 10

consecutive nucleotides of sense or antisense sequence which encodes a gag protein; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence from nucleotides 2452-4839 (e.g, from nucleotides 3112-4683) of SEQ ID NO:1, nucleotides 598-2169 (e.g, from nucleotides 598-2169) of SEQ ID NO:2, or nucleotides 585-2156 (e.g, from
5 nucleotides 585-2156) of SEQ ID NO:3, or naturally occurring mutants thereof;

a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence which encodes a pol protein; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence from nucleotides 4871-8060 of SEQ ID NO:1, nucleotides 2320-4737 of SEQ ID NO:2, or nucleotides 2307-5741 of SEQ ID NO:3, or naturally occurring mutants
10 thereof; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence which encodes a env protein; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence from nucleotides 2-1999 (e.g, from nucleotides 86-1999) of SEQ ID NO:1, nucleotides 4738-6722 (e.g, from nucleotides 4738-6722) of SEQ ID NO:2, or nucleotides 5620-7533 of SEQ ID NO:3, or naturally occurring mutants thereof; a swine or
15 miniature swine retroviral nucleic acid; or a Tsukuba nucleic acid under conditions in which hybridization can occur, hybridization being indicative of the presence or expression of an endogenous miniature swine retrovirus or retroviral sequence or swine retrovirus or retroviral sequence in the genetically modified miniature swine or swine.

In preferred embodiments, the method further includes amplifying the target nucleic
20 acid with primers which specifically hybridize to the sequence of SEQ ID NO:1 or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement.

In preferred embodiments, the target nucleic acid includes: genomic DNA isolated from a miniature swine; RNA or cDNA, e.g., cDNA made from an RNA template, isolated from a miniature swine; DNA, RNA or cDNA, e.g., cDNA made from an RNA template,
25 isolated from a miniature swine organ, e.g., a kidney; RNA, DNA or cDNA, e.g., cDNA made from an RNA template, isolated from a miniature swine potential donor organ; RNA, DNA or cDNA, e.g., cDNA made from an RNA template, isolated from a miniature swine organ which has been transplanted into a organ recipient, e.g., a xenogeneic recipient, e.g., a primate, e.g., a human.

30 In preferred embodiments, the target nucleic acid includes: genomic DNA isolated from a swine; RNA or cDNA, e.g., cDNA made from an RNA template, isolated from a swine; DNA, RNA or cDNA, e.g., cDNA made from an RNA template, isolated from a swine organ, e.g., a kidney; RNA, DNA or cDNA, e.g., cDNA made from an RNA template, isolated from a swine potential donor organ; RNA, DNA or cDNA, e.g., cDNA made from an
35 RNA template, isolated from a swine organ which has been transplanted into a organ recipient, e.g., a recipient swine or a xenogeneic recipient, e.g., a primate, e.g., a human.

In preferred embodiments, the second nucleic acid has at least 60%, 70%, 72%, more preferably at least 85%, more preferably at least 90%, more preferably at least 95%, most preferably at least 98%, 99% or 100% sequence identity or homology with a sequence from

SEQ ID NO:1 or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement.

In other preferred embodiments: the second nucleic acid is at least 10, more preferably at least 15, more preferably at least 20, most preferably at least 25, 30, 50, 100, 1000, 2000, 4000, 6000, or 8060 nucleotides in length; the nucleic acid is less than 15, more preferably less than 20, most preferably less than 25, 30, 50, 100, 1000, 2000, 4000, 6000, or 8060 nucleotides in length; the second nucleic acid is a full length retroviral genome.

In another aspect, the invention features a method of assessing the potential risk associated with the transplantation of a graft from a donor miniature swine or swine into a recipient animal, e.g., a miniature swine or swine, a non-human primate, or a human. The method includes:

contacting a target nucleic acid from the donor, recipient or the graft, with a second sequence chosen from the group of: a nucleic acid sequence which specifically hybridizes a sequence which can specifically hybridize to a porcine retroviral sequence; a sequence which can specifically hybridize to the sequence of SEQ ID NO:1 or its complement; a sequence which can specifically hybridize to the sequence of SEQ ID NO:2 or its complement; a sequence which can specifically hybridize to the sequence of SEQ ID NO:3 or its complement; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence which encodes a gag protein; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence from nucleotides 2452-4839 (e.g, from nucleotides 3112-4683) of SEQ ID NO:1, nucleotides 598-2169 (e.g, from nucleotides 598-2169) of SEQ ID NO:2, or nucleotides 585-2156 (e.g, from nucleotides 585-2156) of SEQ ID NO:3, or naturally occurring mutants thereof;

a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence which encodes a pol protein; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence from nucleotides 4871-8060 of SEQ ID NO:1, nucleotides 2320-4737 of SEQ ID NO:2, or nucleotides 2307-5741 of SEQ ID NO:3, or naturally occurring mutants thereof;

a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence which encodes a env protein; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence from nucleotides 2-1999 (e.g, from nucleotides 86-1999) of SEQ ID NO:1, nucleotides 4738-6722 (e.g, from nucleotides 4738-6722) of SEQ ID NO:2, or nucleotides 5620-7533 of SEQ ID NO:3, or naturally occurring mutants thereof; a swine or miniature swine retroviral nucleic acid; or a Tsukuba nucleic acid under conditions in which the sequences can hybridize, hybridization being indicative of a risk associated with the transplantation.

In a preferred embodiment: the second nucleic acid is a Tsukuba-1 retroviral sequence, probe or primer, e.g., as described herein; the second nucleic acid is a porcine retroviral sequence, probe or primer, e.g., as described herein; the second nucleic acid is the

sequence of SEQ ID NO:1 or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement, or a fragment of the sequence or complement at least 10, 20, or 30, basepairs in length.

5 In preferred embodiments, the target nucleic acid includes: genomic DNA isolated from a miniature swine; RNA or cDNA, e.g., cDNA made from an RNA template, isolated from a miniature swine; DNA, RNA or cDNA, e.g., cDNA made from an RNA template, isolated from a miniature swine organ, e.g., a kidney; RNA, DNA or cDNA, e.g., cDNA made from an RNA template, isolated from a miniature swine potential donor organ; RNA, DNA or cDNA, e.g., cDNA made from an RNA template, isolated from a miniature swine organ which has been transplanted into a organ recipient, e.g., a xenogeneic recipient, e.g., a primate, e.g., a human.

15 In preferred embodiments, the target nucleic acid includes: genomic DNA isolated from a swine; RNA or cDNA, e.g., cDNA made from an RNA template, isolated from a swine; DNA, RNA or cDNA, e.g., cDNA made from an RNA template, isolated from a swine organ, e.g., a kidney; RNA, DNA or cDNA, e.g., cDNA made from an RNA template, isolated from a swine potential donor organ; RNA, DNA or cDNA, e.g., cDNA made from an RNA template, isolated from a swine organ which has been transplanted into a organ recipient, e.g., a recipient swine or a xenogeneic recipient, e.g., a primate, e.g., a human.

20 In preferred embodiments, the second nucleic acid has at least 60%, 70%, 72%, more preferably at least 85%, more preferably at least 90%, more preferably at least 95%, most preferably at least 98%, 99% or 100% sequence identity or homology with a sequence from SEQ ID NO:1 or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement.

25 In other preferred embodiments: the second nucleic acid is at least 10, more preferably at least 15, more preferably at least 20, most preferably at least 25, 30, 50, 100, 1000, 2000, 4000, 6000, or 8060 nucleotides in length; the nucleic acid is less than 15, more preferably less than 20, most preferably less than 25, 30, 50, 100, 1000, 2000, 4000, 6000, or 8060 nucleotides in length; the second nucleic acid is a full length retroviral genome.

30 In another aspect, the invention features a method of determining if an endogenous miniature swine or swine retrovirus or retroviral sequence genome includes a mutation which modulates its expression, e.g., results in misexpression. The method includes:

35 determining the structure of the endogenous retroviral genome, and comparing the structure of the endogenous retroviral genome with the retroviral sequence of SEQ ID NO:1 or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement, a difference being predictive of a mutation.

In preferred embodiments the method includes sequencing the endogenous genome and comparing it with a sequence from SEQ ID NO:1 or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement.

In preferred embodiments, the method includes using primers to amplify, e.g., by PCR, LCR (ligase chain reaction), or other amplification methods, a region of the endogenous retroviral genome, and comparing the structure of the amplification product to the sequence of SEQ ID NO:1 or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement to determine if there is difference in sequence between retroviral genome and SEQ ID NO:1 or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement. The method further includes determining if one or more restriction sites exist in the endogenous retroviral genome, and determining if the sites exist in SEQ ID NO:1 or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement.

In preferred embodiments, the mutation is a gross defect, e.g., an insertion, inversion, translocation or a deletion, of all or part of the retroviral genome.

In preferred embodiments, detecting the mutation can include: (i) providing a labeled PCR probe amplified from DNA (e.g., SEQ ID NO: 1, SEQ ID NO: 2, or SEQ ID NO: 3) containing a porcine retroviral nucleotide sequence which hybridizes to a sense or antisense sequence from the porcine retroviral genome (e.g., SEQ ID NO: 1, SEQ ID NO: 2, or SEQ ID NO: 3), or naturally occurring mutants thereof; (ii) exposing the probe/primer to nucleic acid of the tissue (e.g., genomic DNA) digested with a restriction endonuclease; and (iii) detecting by *in situ* hybridization of the probe/primer to the nucleic acid, the presence or absence of the genetic lesion. Alternatively, direct PCR analysis, using primers specific for porcine retroviral genes (e.g., genes comprising the nucleotide sequence shown in SEQ ID NO: 1, SEQ ID NO: 2, or SEQ ID NO: 3), can be used to detect the presence or absence of the genetic lesion in the porcine retroviral genome by comparing the products amplified.

In another aspect, the invention features a method of providing a miniature swine or a swine free of an endogenous retrovirus or retroviral sequence, e.g., activatable retrovirus, insertion at a preselected site. The method includes:

performing a breeding cross between a first miniature swine (or swine) having a retroviral insertion at the preselected site and a second miniature swine (or swine) not having a retroviral insertion at a preselected site, e.g., the same site, and recovering a progeny miniature swine (or swine), not having the insertion, wherein the presence or absence of the retroviral insertion is determined by contacting the genome of a miniature swine (or swine) with a sequence which can specifically hybridize to a porcine retroviral sequence; a sequence which can specifically hybridize to the sequence of SEQ ID NO:1 or its complement; a sequence which can specifically hybridize to the sequence of SEQ ID NO:2 or its complement; a sequence which can specifically hybridize to the sequence of SEQ ID NO:3 or its complement; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence which encodes a gag protein; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence from nucleotides 2452-4839 (e.g., from nucleotides 3112-4683) of SEQ ID NO:1, nucleotides 598-2169 (e.g., from nucleotides 598-2169) of SEQ ID

NO:2, or nucleotides 585-2156 (e.g, from nucleotides 585-2156) of SEQ ID NO:3, or naturally occurring mutants thereof;

5 a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence which encodes a pol protein; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence from nucleotides 4871-8060 of SEQ ID NO:1, nucleotides 2320-4737 of
10 SEQ ID NO:2, or nucleotides 2307-5741 of SEQ ID NO:3, or naturally occurring mutants thereof; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence which encodes a env protein; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence from nucleotides 2-1999 (e.g, from nucleotides 86-1999) of SEQ ID
15 NO:1, nucleotides 4738-6722 (e.g, from nucleotides 4738-6722) of SEQ ID NO:2, or nucleotides 5620-7533 of SEQ ID NO:3, or naturally occurring mutants thereof; a swine or miniature swine retroviral nucleic acid; or a Tsukuba nucleic acid.

In preferred embodiments, the nucleic acid is hybridized to nucleic acid, e.g., DNA from the genome, of the first animal or one of its ancestors.

15 In preferred embodiments, the nucleic acid is hybridized to nucleic acid, e.g., DNA from the genome, of the second animal or one of its ancestors.

In preferred embodiments, the nucleic acid is hybridized to nucleic acid, e.g., DNA from the genome, of the progeny animal or one of its descendants.

20 In preferred embodiments, the nucleic acid has at least 60%, 70%, 72%, more preferably at least 85%, more preferably at least 90%, more preferably at least 95%, most preferably at least 98%, 99% or 100% sequence identity or homology with a sequence from SEQ ID NO:1 or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement.

25 In other preferred embodiments: the nucleic acid is at least 10, more preferably at least 15, more preferably at least 20, most preferably at least 25, 30, 50, 100, 1000, 2000, 4000, 6000, or 8060 nucleotides in length; the nucleic acid is less than 15, more preferably less than 20, most preferably less than 25, 30, 50, 100, 1000, 2000, 4000, 6000, or 8060 nucleotides in length; the nucleic acid is a full length retroviral genome.

30 In another aspect, the invention features a method of evaluating a treatment, e.g., an immunosuppressive treatment, for the ability to activate a retrovirus, e.g., an endogenous porcine retrovirus. The method includes:

administering a treatment to a subject, e.g., a miniature swine (or a swine), having an endogenous porcine retrovirus; and

35 detecting expression of the porcine retrovirus with a purified nucleic acid sequence which specifically hybridizes to the sequence of SEQ ID NO:1 or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement.

In preferred embodiments, the immunosuppressive treatment includes radiation, chemotherapy or drug treatment.

In preferred embodiments: the treatment is one which can induce immunological tolerance; the treatment is one which can introduce new genetic material, e.g., introduce new genetic material into a miniature swine genome (or a swine genome) or into the genome of a host which receives a swine or a miniature swine graft, e.g., the treatment is one which introduces a new genetic material via retroviral mediated transfer.

In a preferred embodiment: the purified nucleic acid is a Tsukuba-1 retroviral sequence, probe or primer, e.g., as described herein; the purified nucleic acid is a porcine retroviral sequence, probe or primer, e.g., as described herein; the purified nucleic acid is the sequence of SEQ ID NO:1 or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement, or a fragment of such sequence or complement at least 10, 20, or 30, basepairs in length.

In preferred embodiments, the purified nucleic acid has at least 60%, 70%, 72%, more preferably at least 85%, more preferably at least 90%, more preferably at least 95%, most preferably at least 98%, 99% or 100% sequence identity or homology with a sequence from SEQ ID NO:1 or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement.

In other preferred embodiments: the purified nucleic acid is at least 10, more preferably at least 15, more preferably at least 20, most preferably at least 25, 30, 50, 100, 1000, 2000, 4000, 6000, or 8060 nucleotides in length; the nucleic acid is less than 15, more preferably less than 20, most preferably less than 25, 30, 50, 100, 1000, 2000, 4000, 6000, or 8060 nucleotides in length; the purified nucleic acid is a full length retroviral genome.

In preferred embodiments the second nucleic acid is: a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence which encodes a gag protein; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence from nucleotides 2452-4839 (e.g, from nucleotides 3112-4683) of SEQ ID NO:1, nucleotides 598-2169 (e.g, from nucleotides 598-2169) of SEQ ID NO:2, or nucleotides 585-2156 (e.g, from nucleotides 585-2156) of SEQ ID NO:3, or naturally occurring mutants thereof; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence which encodes a pol protein; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence from nucleotides 4871-8060 of SEQ ID NO:1, nucleotides 2320-4737 of SEQ ID NO:2, or nucleotides 2307-5741 of SEQ ID NO:3, or naturally occurring mutants thereof; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence which encodes a env protein; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence from nucleotides 2-1999 (e.g, from nucleotides 86-1999) of SEQ ID NO:1, nucleotides 4738-6722 (e.g, from nucleotides 4738-6722) of SEQ ID NO:2, or nucleotides 5620-7533 of SEQ ID NO:3, or naturally occurring mutants thereof.

In another aspect, the invention features a method of localizing the origin of a porcine retroviral infection. The method includes:

contacting a target nucleic acid from the graft with a second sequence chosen from the group of: a sequence which can specifically hybridize to a porcine retroviral sequence; a sequence which can specifically hybridize to the sequence of SEQ ID NO:1 or its complement; a sequence which can specifically hybridize to the sequence of SEQ ID NO:2 or its complement; a sequence which can specifically hybridize to the sequence of SEQ ID NO:3 or its complement; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence which encodes a gag protein; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence from nucleotides 2452-4839 (e.g, from nucleotides 3112-4683) of SEQ ID NO:1, nucleotides 598-2169 (e.g, from nucleotides 598-2169) of SEQ ID NO:2, or nucleotides 585-2156 (e.g, from nucleotides 585-2156) of SEQ ID NO:3, or naturally occurring mutants thereof; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence which encodes a pol protein; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence from nucleotides 4871-8060 of SEQ ID NO:1, nucleotides 2320-4737 of SEQ ID NO:2, or nucleotides 2307-5741 of SEQ ID NO:3, or naturally occurring mutants thereof;

a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence which encodes a env protein; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence from nucleotides 2-1999 (e.g, from nucleotides 86-1999) of SEQ ID NO:1, nucleotides 4738-6722 (e.g, from nucleotides 4738-6722) of SEQ ID NO:2, or nucleotides 5620-7533 of SEQ ID NO:3, or naturally occurring mutants thereof; a swine or miniature swine retroviral nucleic acid; or a Tsukuba nucleic acid contacting a target nucleic acid from the recipient with a second sequence chosen from the group of: a sequence which can specifically hybridize to a porcine retroviral sequence; a sequence which can specifically hybridize to the sequence of SEQ ID NO:1 or its complement; a sequence which can specifically hybridize to the sequence of SEQ ID NO:2 or its complement; a sequence which can specifically hybridize to the sequence of SEQ ID NO:3 or its complement; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence which encodes a gag protein; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence from nucleotides 2452-4839 (e.g, from nucleotides 3112-4683) of SEQ ID NO:1, nucleotides 598-2169 (e.g, from nucleotides 598-2169) of SEQ ID NO:2, or nucleotides 585-2156 (e.g, from nucleotides 585-2156) of SEQ ID NO:3, or naturally occurring mutants thereof; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence which encodes a pol protein; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence from nucleotides 4871-8060 of SEQ ID NO:1, nucleotides 2320-4737 of SEQ ID NO:2, or nucleotides 2307-5741 of SEQ ID NO:3, or naturally occurring mutants thereof; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence which encodes a env protein; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence from nucleotides 2-1999 (e.g, from nucleotides 86-1999) of SEQ ID NO:1, nucleotides 4738-6722 (e.g, from nucleotides 4738-6722) of SEQ ID NO:2, or

nucleotides 5620-7533 of SEQ ID NO:3, or naturally occurring mutants thereof; a swine or miniature swine retroviral nucleic acid; or a Tsukuba nucleic acid; hybridization to the nucleic acid from the graft correlates with the porcine retroviral infection in the graft; and hybridization to the nucleic acid from the recipient correlates with the porcine retroviral infection in the recipient.

In preferred embodiments, the target nucleic acid includes: genomic DNA, RNA or cDNA, e.g., cDNA made from an RNA template.

In a preferred embodiment: the second nucleic acid is a porcine retroviral sequence, probe or primer, e.g., as described herein, e.g., a Tsukuba-1 retroviral sequence; the second nucleic acid is a sequence of SEQ ID NO:1 or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement, or a fragment of the sequence or complement at least 10, 20, or 30, basepairs in length.

In preferred embodiments, the recipient is an animal, e.g., a miniature swine, a swine, a non-human primate, or a human.

In preferred embodiments, the graft is selected from the group consisting of: heart, lung, liver, bone marrow or kidney.

In preferred embodiments, the second nucleic acid has at least 60%, 70%, 72%, more preferably at least 85%, more preferably at least 90%, more preferably at least 95%, most preferably at least 98%, 99% or 100% sequence identity or homology with a sequence from SEQ ID NO:1 or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement.

In other preferred embodiments: the second nucleic acid is at least 10, more preferably at least 15, more preferably at least 20, most preferably at least 25, 30, 50, 100, 1000, 2000, 4000, 6000, or 8060 nucleotides in length; the nucleic acid is less than 15, more preferably less than 20, most preferably less than 25, 30, 50, 100, 1000, 2000, 4000, 6000, or 8060 nucleotides in length; the second nucleic acid is a full length retroviral genome.

In another aspect, the invention features a method of screening a cell, e.g., a cell having a disorder, e.g., a proliferative disorder, e.g., a tumor cell, e.g., a cancer cell, e.g., a lymphoma or a hepatocellular carcinoma, developing in a graft recipient, e.g., a xenograft, for the presence or expression of a porcine retrovirus or retroviral sequence. The method includes:

contacting a target nucleic acid from a tumor cell with a second sequence chosen from the group of: a sequence which can specifically hybridize to a porcine retroviral sequence; a sequence which can specifically hybridize to the sequence of SEQ ID NO:1 or its complement; a sequence which can specifically hybridize to the sequence of SEQ ID NO:2 or its complement; a sequence which can specifically hybridize to the sequence of SEQ ID NO:3 or its complement; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence which encodes a gag protein; a nucleic acid of at least 10 consecutive

nucleotides of sense or antisense sequence from nucleotides 2452-4839 (e.g, from nucleotides 3112-4683) of SEQ ID NO:1, nucleotides 598-2169 (e.g, from nucleotides 598-2169) of SEQ ID NO:2, or nucleotides 585-2156 (e.g, from nucleotides 585-2156) of SEQ ID NO:3, or naturally occurring mutants thereof; a nucleic acid of at least 10 consecutive
5 nucleotides of sense or antisense sequence which encodes a pol protein; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence from nucleotides 4871-8060 of SEQ ID NO:1, nucleotides 2320-4737 of SEQ ID NO:2, or nucleotides 2307-5741 of SEQ ID NO:3, or naturally occurring mutants thereof; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence which encodes a env protein; a
10 nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence from nucleotides 2-1999 (e.g, from nucleotides 86-1999) of SEQ ID NO:1, nucleotides 4738-6722 (e.g, from nucleotides 4738-6722) of SEQ ID NO:2, or nucleotides 5620-7533 of SEQ ID NO:3, or naturally occurring mutants thereof; a swine or miniature swine retroviral nucleic acid; or a Tsukuba nucleic acid, under conditions in which the sample and the nucleic acid
15 sequence can hybridize, hybridization being indicative of the presence of the endogenous porcine retrovirus or retroviral sequence in the tumor cell.

In preferred embodiments, the target nucleic acid from a tumor cell includes: genomic DNA, RNA or cDNA, e.g., cDNA made from an RNA template.

In a preferred embodiment: the second nucleic acid is a porcine retroviral sequence, probe or primer, e.g., as described herein, e.g., a Tsukuba-1 retroviral sequence; the second
20 nucleic acid is a sequence of SEQ ID NO:1 or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement, or a fragment of the sequence or complement at least 10, 20, or 30, basepairs in length.

In preferred embodiments, the second nucleic acid has at least 60%, 70%, 72%, more preferably at least 85%, more preferably at least 90%, more preferably at least 95%, most preferably at least 98%, 99% or 100% sequence identity or homology with a sequence from
25 SEQ ID NO:1 or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement.

In other preferred embodiments: the second nucleic acid is at least 10, more preferably at least 15, more preferably at least 20, most preferably at least 25, 30, 50, 100, 1000, 2000, 4000, 6000, or 8060 nucleotides in length; the nucleic acid is less than 15, more preferably less than 20, most preferably less than 25, 30, 50, 100, 1000, 2000, 4000, 6000, or 8060 nucleotides in length; the second nucleic acid is a full length retroviral genome.

In another aspect, the invention features a method of screening a human subject for
35 the presence or expression of an endogenous porcine retrovirus or retroviral sequence comprising:

contacting a target nucleic acid derived from the human subject with a second sequence chosen from the group of: a sequence which can specifically hybridize to a porcine retroviral sequence; a sequence which can specifically hybridize to the sequence of SEQ ID

NO:1 or its complement; a sequence which can specifically hybridize to the sequence of SEQ ID NO:2 or its complement; a sequence which can specifically hybridize to the sequence of SEQ ID NO:3 or its complement; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence which encodes a gag protein; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence from nucleotides 2452-4839 (e.g, 5 from nucleotides 3112-4683) of SEQ ID NO:1, nucleotides 598-2169 (e.g, from nucleotides 598-2169) of SEQ ID NO:2, or nucleotides 585-2156 (e.g, from nucleotides 585-2156) of SEQ ID NO:3, or naturally occurring mutants thereof; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence which encodes a pol protein; a 10 nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence from nucleotides 4871-8060 of SEQ ID NO:1, nucleotides 2320-4737 of SEQ ID NO:2, or nucleotides 2307-5741 of SEQ ID NO:3, or naturally occurring mutants thereof; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence which encodes a env protein; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense 15 sequence from nucleotides 2-1999 (e.g, from nucleotides 86-1999) of SEQ ID NO:1, nucleotides 4738-6722 (e.g, from nucleotides 4738-6722) of SEQ ID NO:2, or nucleotides 5620-7533 of SEQ ID NO:3, or naturally occurring mutants thereof; a swine or miniature swine retroviral nucleic acid; or a Tsukuba nucleic acid under conditions in which the sequences can hybridize, hybridization being indicative of the presence of the endogenous 20 porcine retrovirus or retroviral sequence in the human subject.

In preferred embodiments, the target nucleic acid derived from a human subject is DNA, RNA or cDNA sample, nucleic acid from a blood sample or a tissue sample, e.g., a tissue biopsy sample.

In preferred embodiments, the human subject is a miniature swine or swine xenograft 25 recipient, or a person who has come into contact with a miniature swine or swine xenograft recipient.

In preferred embodiments, the second nucleic acid has at least 60%, 70%, 72%, more preferably at least 85%, more preferably at least 90%, more preferably at least 95%, most preferably at least 98%, 99% or 100% sequence identity or homology with a sequence from 30 SEQ ID NO:1 or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement.

In other preferred embodiments: the second nucleic acid is at least 10, more preferably at least 15, more preferably at least 20, most preferably at least 25, 30, 50, 100, 1000, 2000, 4000, 6000, or 8060 nucleotides in length; the nucleic acid is less than 15, more 35 preferably less than 20, most preferably less than 25, 30, 50, 100, 1000, 2000, 4000, 6000, or 8060 nucleotides in length; the second nucleic acid is a full length retroviral genome.

In preferred embodiments: the recipient is tested for the presence of porcine retroviral sequences prior to implantation of swine or miniature swine tissue.

In another aspect, the invention features a method of screening for viral mutations which modulate, e.g., increase or decrease, susceptibility of a porcine retrovirus to an antiviral agent, e.g., an antiviral antibiotic. The method includes:

- administering a treatment, e.g., an antiviral agent, e.g., an antiviral antibiotic;
- 5 isolating a putative mutant porcine retroviral strain;
- determining a structure of the putative mutant retroviral strain; and
- comparing the structure to SEQ ID NO:1 or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement.

In another aspect, the invention features a method of screening for viral mutations which modulate, e.g., increase or decrease, susceptibility of a porcine retrovirus to an antiviral agent, e.g., an antiviral antibiotic. The method includes:

- growing the porcine retrovirus in a presence of a treatment, e.g., an antiviral agent, e.g., an antiviral antibiotic; and
- determine the amount of porcine retroviral DNA synthesized by hybridizing the
- 15 porcine retroviral DNA to a second sequence chosen from the group of: a sequence which can specifically hybridize to a porcine retroviral sequence; a sequence which can specifically hybridize to the sequence of SEQ ID NO:1 or its complement; a sequence which can specifically hybridize to the sequence of SEQ ID NO:2 or its complement; a sequence which can specifically hybridize to the sequence of SEQ ID NO:3 or its complement; a nucleic acid
- 20 of at least 10 consecutive nucleotides of sense or antisense sequence which encodes a gag protein; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence from nucleotides 2452-4839 (e.g, from nucleotides 3112-4683) of SEQ ID NO:1, nucleotides 598-2169 (e.g, from nucleotides 598-2169) of SEQ ID NO:2, or nucleotides 585-2156 (e.g, from nucleotides 585-2156) of SEQ ID NO:3, or naturally occurring mutants thereof; a
- 25 nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence which encodes a pol protein; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence from nucleotides 4871-8060 of SEQ ID NO:1, nucleotides 2320-4737 of SEQ ID NO:2, or nucleotides 2307-5741 of SEQ ID NO:3, or naturally occurring mutants thereof; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence
- 30 which encodes a env protein; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence from nucleotides 2-1999 (e.g, from nucleotides 86-1999) of SEQ ID NO:1, nucleotides 4738-6722 (e.g, from nucleotides 4738-6722) of SEQ ID NO:2, or nucleotides 5620-7533 of SEQ ID NO:3, or naturally occurring mutants thereof; a swine or miniature swine retroviral nucleic acid; or a Tsukuba nucleic acid.

35 In preferred embodiments, the method further includes amplifying the porcine retroviral nucleic acid with primers which specifically hybridize to the sequence of SEQ ID NO:1 or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement, e.g., by polymerase chain reaction quantitative DNA testing (PDQ).

In a preferred embodiment: the second nucleic acid is a Tsukuba-1 retroviral sequence, probe or primer, e.g., as described herein; the second nucleic acid is a porcine retroviral sequence, probe or primer, e.g., as described herein; the second nucleic acid is the sequence of SEQ ID NO:1 or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement.

In preferred embodiments, the second nucleic acid has at least 60%, 70%, 72%, more preferably at least 85%, more preferably at least 90%, more preferably at least 95%, most preferably at least 98%, 99% or 100% sequence identity or homology with a sequence from SEQ ID NO:1 or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement.

In other preferred embodiments: the second nucleic acid is at least 10, more preferably at least 15, more preferably at least 20, most preferably at least 25, 30, 50, 100, 1000, 2000, 4000, 6000, or 8060 nucleotides in length; the nucleic acid is less than 15, more preferably less than 20, most preferably less than 25, 30, 50, 100, 1000, 2000, 4000, 6000, or 8060 nucleotides in length; the second nucleic acid is a full length retroviral genome.

In another aspect, the invention features a method for screening a porcine-derived product for the presence or expression of a swine or miniature swine retrovirus or retroviral sequence, e.g., an endogenous miniature swine retrovirus. The method includes:

contacting a target nucleic acid from the porcine-derived product with a second sequence chosen from the group of: a sequence which can specifically hybridize to a porcine retroviral sequence; a sequence which can specifically hybridize to the sequence of SEQ ID NO:1 or its complement; a sequence which can specifically hybridize to the sequence of SEQ ID NO:2 or its complement; a sequence which can specifically hybridize to the sequence of SEQ ID NO:3 or its complement; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence which encodes a gag protein; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence from nucleotides 2452-4839 (e.g., from nucleotides 3112-4683) of SEQ ID NO:1, nucleotides 598-2169 (e.g., from nucleotides 598-2169) of SEQ ID NO:2, or nucleotides 585-2156 (e.g., from nucleotides 585-2156) of SEQ ID NO:3, or naturally occurring mutants thereof; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence which encodes a pol protein; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence from nucleotides 4871-8060 of SEQ ID NO:1, nucleotides 2320-4737 of SEQ ID NO:2, or nucleotides 2307-5741 of SEQ ID NO:3, or naturally occurring mutants thereof; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence which encodes a env protein; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence from nucleotides 2-1999 (e.g., from nucleotides 86-1999) of SEQ ID NO:1, nucleotides 4738-6722 (e.g., from nucleotides 4738-6722) of SEQ ID NO:2, or nucleotides 5620-7533 of SEQ ID NO:3, or naturally occurring mutants thereof; a swine or miniature swine retroviral nucleic acid; or a Tsukuba nucleic acid, under conditions in which

hybridization can occur, hybridization being indicative of the presence or expression of an endogenous miniature swine or swine retrovirus or retroviral sequence s in the porcine-derived product.

5 In preferred embodiments the product is: a protein product, e.g., insulin; a food product; or a cellular transplant, e.g., a swine or miniature swine cell which is to be transplanted into a host, e.g., a swine or miniature swine cell which is genetically engineered to express a desired product,

10 In preferred embodiments, the method further includes amplifying the target nucleic acid with primers which specifically hybridize to the sequence of SEQ ID NO:1 or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement.

In other preferred embodiments, the target nucleic acid is: DNA; RNA; or cDNA.

15 In preferred embodiments, the second nucleic acid has at least 60%, 70%, 72%, more preferably at least 85%, more preferably at least 90%, more preferably at least 95%, most preferably at least 98%, 99% or 100% sequence identity or homology with a sequence from SEQ ID NO:1 or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement.

20 In other preferred embodiments: the second nucleic acid is at least 10, more preferably at least 15, more preferably at least 20, most preferably at least 25, 30, 50, 100, 1000, 2000, 4000, 6000, or 8060 nucleotides in length; the nucleic acid is less than 15, more preferably less than 20, most preferably less than 25, 30, 50, 100, 1000, 2000, 4000, 6000, or 8060 nucleotides in length; the second nucleic acid is a full length retroviral genome.

25 In another aspect, the invention features a transgenic miniature swine or swine having a transgenic element, e.g., a base change, e.g., a change from A to G, or an insertion or a deletion of one or more nucleotides at an endogenous porcine retroviral insertion site, e.g., a retroviral insertion which corresponds to the retroviral genome of SEQ ID NO:1 or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement.

In preferred embodiments, the transgenic element is a knockout, e.g., a deletion, insertion or a translocation, of one or more nucleic acids, which alters the activity of the endogenous porcine retrovirus.

30 In another aspect, the invention features a method of inhibiting expression of an endogenous porcine retrovirus, including: inserting a mutation, e.g. a deletion into the endogenous retrovirus.

In preferred embodiments, the endogenous porcine retrovirus is inactivated.

35 In preferred embodiments, the mutation can be a point mutation, an inversion, translocation or a deletion of one or more nucleotides of SEQ ID NO:1 or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement.

In another aspect, the invention features a method of detecting a recombinant virus or other pathogen, e.g., a protozoa or fungi. The method includes:

providing a pathogen having porcine retroviral sequence; and

determining if the pathogen includes non-porcine retroviral sequence, the presence of non-porcine retroviral sequence being indicative of viral recombination.

In preferred embodiments, the method further includes determining the structure of a retrovirus by comparing the retrovirus sequence with sequence of SEQ ID NO:1 or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement, a difference being indicative of viral recombination.

In preferred embodiments, the method further includes comparing the structure of the retrovirus with a human retroviral sequence, e.g., HTLV1, HIV1, or HIV2, a similarity in structure being indicative of viral recombination.

In another aspect, the invention features a method of determining the copy number, size, or completeness of a porcine retrovirus or retroviral sequence, e.g., in the genome of a donor, recipient or a graft. The method includes:

contacting a target nucleic acid from the donor, recipient or a graft, with a second sequence chosen from the group of: a sequence which can specifically hybridize to a porcine retroviral sequence; a sequence which can specifically hybridize to the sequence of SEQ ID NO:1 or its complement; a sequence which can specifically hybridize to the sequence of SEQ ID NO:2 or its complement; a sequence which can specifically hybridize to the sequence of SEQ ID NO:3 or its complement; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence which encodes a gag protein; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence from nucleotides 2452-4839 (e.g, from nucleotides 3112-4683) of SEQ ID NO:1, nucleotides 598-2169 (e.g, from nucleotides 598-2169) of SEQ ID NO:2, or nucleotides 585-2156 (e.g, from nucleotides 585-2156) of SEQ ID NO:3, or naturally occurring mutants thereof;

a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence which encodes a pol protein; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence from nucleotides 4871-8060 of SEQ ID NO:1, nucleotides 2320-4737 of SEQ ID NO:2, or nucleotides 2307-5741 of SEQ ID NO:3, or naturally occurring mutants thereof;

a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence which encodes a env protein; a nucleic acid of at least 10 consecutive nucleotides of sense or antisense sequence from nucleotides 2-1999 (e.g, from nucleotides 86-1999) of SEQ ID NO:1, nucleotides 4738-6722 (e.g, from nucleotides 4738-6722) of SEQ ID NO:2, or nucleotides 5620-7533 of SEQ ID NO:3, or naturally occurring mutants thereof; a swine or miniature swine retroviral nucleic acid; or a Tsukuba nucleic acid.

In preferred embodiments, the method further includes amplifying the porcine retroviral nucleic acid with primers which specifically hybridize to the sequence of SEQ ID NO:1 or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement, e.g., by polymerase chain reaction quantitative DNA testing (PDQ) or nested PCR.

In preferred embodiments, the target nucleic acid includes: genomic DNA isolated from a miniature swine; RNA or cDNA, e.g., cDNA made from an RNA template, isolated from a miniature swine; DNA, RNA or cDNA, e.g., cDNA made from an RNA template, isolated from a miniature swine organ, e.g., a kidney; RNA, DNA or cDNA, e.g., cDNA made from an RNA template, isolated from a miniature swine organ which has been transplanted into a organ recipient, e.g., a xenogeneic recipient, e.g., a primate, e.g., a human.

In preferred embodiments, the target nucleic acid includes: genomic DNA isolated from a swine; RNA or cDNA, e.g., cDNA made from an RNA template, isolated from a swine; DNA, RNA or cDNA, e.g., cDNA made from an RNA template, isolated from a swine organ, e.g., a kidney; RNA, DNA or cDNA, e.g., cDNA made from an RNA template, isolated from a swine organ which has been transplanted into a organ recipient, e.g., a xenogeneic recipient, e.g., a primate, e.g., a human.

In preferred embodiments, the second nucleic acid has at least 60%, 70%, 72%, more preferably at least 85%, more preferably at least 90%, more preferably at least 95%, most preferably at least 98%, 99% or 100% sequence identity or homology with a sequence from SEQ ID NO:1 or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement.

In other preferred embodiments: the second nucleic acid is at least 10, more preferably at least 15, more preferably at least 20, most preferably at least 25, 30, 50, 100, 1000, 2000, 4000, 6000, or 8060 nucleotides in length; the nucleic acid is less than 15, more preferably less than 20, most preferably less than 25, 30, 50, 100, 1000, 2000, 4000, 6000, or 8060 nucleotides in length; the second nucleic acid is a full length retroviral genome.

In another aspect, the invention features a method for screening a tissue, e.g., a cellular or tissue transplant, e.g., a xenograft, or a tissue from a graft recipient, for the presence or expression of a swine or a miniature swine retroviral sequence, e.g., an endogenous miniature swine retrovirus. The method includes: contacting a tissue sample with an antibody specific for a retroviral protein, e.g., an anti-gag, pol, or env antibody, and thereby determining if the sequence is present or expressed.

In preferred embodiments the protein is encoded by a sequence from: the sequence of SEQ ID NO:1 or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement.

In preferred embodiments, the tissue is selected from the group consisting of: heart, lung, liver, bone marrow, kidney, brain cells, neural tissue, pancreas or pancreatic cells, thymus, or intestinal tissue.

A "purified preparation" or a "substantially pure preparation" of a polypeptide as used herein, means a polypeptide which is free from one or more other proteins, lipids, and nucleic acids with which it naturally occurs. Preferably, the polypeptide, is also separated from substances which are used to purify it, e.g., antibodies or gel matrix, such as polyacrylamide. Preferably, the polypeptide constitutes at least 10, 20, 50 70, 80 or 95% dry weight of the

purified preparation. Preferably, the preparation contains: sufficient polypeptide to allow protein sequencing; at least 1, 10, or 100 µg of the polypeptide; at least 1, 10, or 100 mg of the polypeptide.

Specifically hybridize, as used herein, means that a nucleic acid hybridizes to a target sequence with substantially greater degree than it does to other sequences in a reaction mixture. By substantially greater means a difference sufficient to determine if the target sequence is present in the mixture.

A "treatment", as used herein, includes any therapeutic treatment, e.g., the administration of a therapeutic agent or substance, e.g., a drug or irradiation.

A "purified preparation of nucleic acid", is a nucleic acid which is one or both of: not immediately contiguous with one or both of the coding sequences with which it is immediately contiguous (i.e., one at the 5' end and one at the 3' end) in the naturally-occurring genome of the organism from which the nucleic acid is derived; or which is substantially free of a nucleic acid sequence or protein with which it occurs in the organism from which the nucleic acid is derived. The term includes, for example, a recombinant DNA which is incorporated into a vector, e.g., into an autonomously replicating plasmid or virus, or into the genomic DNA of a prokaryote or eukaryote, or which exists as a separate molecule (e.g., a cDNA or a genomic DNA fragment produced by PCR or restriction endonuclease treatment) independent of other DNA sequences. Substantially pure DNA also includes a recombinant DNA which is part of a hybrid gene encoding additional sequences. A purified retroviral genome is a nucleic acid which is substantially free of host nucleic acid or viral protein.

"Homologous", as used herein, refers to the sequence similarity between two polypeptide molecules or between two nucleic acid molecules. When a position in both of the two compared sequences is occupied by the same amino acid or base monomer subunit, e.g., if a position in each of two DNA molecules is occupied by adenine, then the molecules are homologous at that position. The percent of homology between two sequences is a function of the number of matching or homologous positions shared by the two sequences divided by the number of positions compared x 100. For example, if 6 of 10, of the positions in two sequences are matched or homologous then the two sequences are 60% homologous. By way of example, the DNA sequences ATTGCC and TATGGC share 50% homology. Generally, a comparison is made when two sequences are aligned to give maximum homology. The term sequence identity has substantially the same meaning.

The term "provirus" or "endogenous retrovirus," as used herein, refers to an integrated form of the retrovirus.

The terms "peptides", "proteins", and "polypeptides" are used interchangeably herein.

As used herein, the term "transgenic element" means a nucleic acid sequence, which is partly or entirely heterologous, i.e., foreign, to the animal or cell into which it is introduced but which is designed to be inserted, or is inserted, into the animal's genome in

such a way as to alter the genome of the cell into which it is inserted. The term includes elements which cause a change in the sequence, or in the ability to be activated, of an endogenous retroviral sequence. Examples of transgenic elements include those which result in changes, e.g., substitutions (e.g., A for G), insertions or deletions of an endogenous retroviral sequence (or flanking regions) which result in inhibition of activation or misexpression of a retroviral product.

As used herein, the term "transgenic cell" refers to a cell containing a transgenic element.

As used herein, a "transgenic animal" is any animal in which one or more, and preferably essentially all, of the cells of the animal includes a transgenic element. The transgenic element can be introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection. This molecule may be integrated within a chromosome, or it may be extrachromosomally replicating DNA.

As described herein, one aspect of the invention features a pure (or recombinant) nucleic acid which includes a miniature swine (or swine) retroviral genome or fragment thereof, e.g., nucleotide sequence encoding a gag-pol or env polypeptide, and/or equivalents of such nucleic acids. The term "nucleic acid", as used herein, can include fragments and equivalents. The term "equivalent" refers to nucleotide sequences encoding functionally equivalent polypeptides or functionally equivalent polypeptides which, for example, retain the ability to react with an antibody specific for a gag-pol or env polypeptide. Equivalent nucleotide sequences will include sequences that differ by one or more nucleotide substitutions, additions or deletions, such as allelic variants, and will, therefore, include sequences that differ from the nucleotide sequence of gag, pol, or env shown in herein due to the degeneracy of the genetic code.

"Misexpression", as used herein, refers to a non-wild type pattern of gene expression, e.g., porcine retroviral, e.g., Tsukuba-1 gene expression, e.g., gag, pol or env gene expression. It includes: expression at non-wild type levels, i.e., over or under expression; a pattern of expression that differs from wild type in terms of the time or stage at which the gene is expressed, e.g., increased or decreased expression (as compared with wild type) at a predetermined developmental period or stage; a pattern of expression that differs from wild type in terms of decreased expression (as compared with wild type) in a predetermined cell type or tissue type; a pattern of expression that differs from wild type in terms of the splicing, size, amino acid sequence, post-translational modification, stability, or biological activity of the expressed porcine retroviral, e.g., Tsukuba-1, polypeptides; a pattern of expression that differs from wild type in terms of the effect of an environmental stimulus or extracellular stimulus on expression of the porcine retroviral, e.g., Tsukuba-1 genes, e.g., a pattern of increased or decreased expression (as compared with wild type) in the presence of an increase or decrease in the strength of the stimulus.

Methods of the invention can be used with swine or miniature swine.

Endogenous retrovirus is a potential source of infection not always susceptible to conventional breeding practices. Many proviruses are defective and unable to replicate. Provirus, if intact, can be activated by certain stimuli and then initiate viral replication using the host's cellular mechanisms. Retroviral infection will often not harm the host cell. However, replication of virus may result in viremia, malignant transformation (e.g., via insertion of retroviral oncogenes), degeneration, or other insertional effects (e.g., gene inactivation). The effects of such infection may not emerge for many years. The spectrum of behavior of active lentiviral infection in humans is well described relative to HIV. These include AIDS, unusual infections and tumors, recombinant and other viruses, and antigenic variation which may prevent the generation of protective immunity by the infected host.

Screening of animals will allow elimination of donors with active replication of known viruses. Inactive proviruses can be detected with genetic probes and removed or inactivated. These novel approaches will allow the identification and elimination of potential human pathogens derived from swine in a manner not possible in the outbred human organ donor population and, thus, will be important to the development of human xenotransplantation.

The porcine retroviral sequences of the invention are also useful as diagnostic probes to detect activation of endogenous porcine retroviruses following transplantation and xenotransplantation of organs derived from swine or miniature swine. The porcine retroviral sequences of the invention also provide diagnostic tools necessary to assess the risks associated with transplantation of organs from swine or miniature swine into human recipients. These sequences are also useful for the longitudinal evaluation of retroviral activation in the human recipient of miniature swine-derived organs.

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of cell biology, cell culture, molecular biology, transgenic biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are described in the literature. See, for example, *Molecular Cloning A Laboratory Manual*, 2nd Ed., ed. by Sambrook, Fritsch and Maniatis (Cold Spring Harbor Laboratory Press: 1989); *DNA Cloning*, Volumes I and II (D. N. Glover ed., 1985); *Oligonucleotide Synthesis* (M. J. Gait ed., 1984); Mullis et al. U.S. Patent No: 4,683,195; *Nucleic Acid Hybridization* (B. D. Hames & S. J. Higgins eds. 1984); *Transcription And Translation* (B. D. Hames & S. J. Higgins eds. 1984); *Culture Of Animal Cells* (R. I. Freshney, Alan R. Liss, Inc., 1987); *Immobilized Cells And Enzymes* (IRL Press, 1986); B. Perbal, *A Practical Guide To Molecular Cloning* (1984); the treatise, *Methods In Enzymology* (Academic Press, Inc., N.Y.); *Gene Transfer Vectors For Mammalian Cells* (J. H. Miller and M. P. Calos eds., 1987, Cold Spring Harbor Laboratory); *Methods In Enzymology*, Vols. 154 and 155 (Wu et al. eds.), *Immunochemical Methods In Cell And Molecular Biology* (Mayer and Walker, eds., Academic Press, London, 1987); *Handbook Of Experimental Immunology*,

Volumes I-IV (D. M. Weir and C. C. Blackwell, eds., 1986); *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986).

Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods and materials are described below. All publications mentioned herein are incorporated by reference. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

Detailed Description of the Drawings

Figure 1 is the nucleotide sequence (SEQ ID NO: 1) of the Tsukuba-1 cDNA.

Figure 2 is the nucleotide sequence (SEQ ID NO: 2) of a defective retroviral genome isolated from the retrovirus from the PK-15 cell line.

Figure 3 is the nucleotide sequence (SEQ ID NO: 3) of a retrovirus found in miniature swine.

Detailed Description

Miniature Swine Retroviruses

Transplantation may increase the likelihood of retroviral activation, if intact and infectious proviruses are present. Many phenomena associated with transplantation, e.g., immune suppression, graft rejection, graft-versus-host disease, viral co-infection, cytotoxic therapies, radiation therapy or drug treatment, can promote activation of retroviral expression.

Many species are thought to carry retroviral sequences in their genomic DNA. The number of intact (complete) retroviral elements that could be activated is often unknown. Once activated, swine-derived viruses would require the appropriate receptor on human tissues to spread beyond the transplanted organ. Most intact endogenous proviruses (usually types B and C), once activated, are not pathogenic. However, coinfection with other viruses, recombination with other endogenous viruses, or modification of viral behavior in the foreign human environment may alter the pathogenicity, organ specificity or replication of the retroviruses or other infectious agents.

The lack of sequence data on pig viruses has impeded efforts to assess the number of porcine sequences, or porcine retroviral sequences, that have incorporated into the human genome or the frequency of incorporation.

The inventor, by showing that the Tsukuba-1 retrovirus is found in miniature swine, and by providing the entire sequence of the porcine retroviral (Tsukuba-1) genome, has allowed assessment of the risk of endogenous retroviruses in general clinical practice and more importantly in xenotransplantation.

The porcine retroviral sequences of the invention can be used to determine the level (e.g., copy number) of intact (i.e., potentially replicating) porcine provirus sequences in a strain of xenograft transplantation donors. For example, the copy number of the miniature swine retroviral sequences can be determined by the Polymerase Chain Reaction DNA Quantitation (PDQ) method, described herein, or by other methods known to those skilled in

the art. This quantitation technique will allow for the selection of animal donors, e.g., miniature swine donors, without an intact porcine retroviral sequence or with a lower copy number of viral elements.

5 The porcine retroviral sequences of the invention can be used to determine if mutations, e.g., inversions, translocations, insertions or deletions, have occurred in the endogenous porcine retroviral sequence. Mutated viral genomes may be expression-deficient. For example, genetic lesions can be identified by exposing a probe/primer derived from porcine retrovirus sequence to nucleic acid of the tissue (e.g., genomic DNA) digested with a restriction endonucleases or by *in situ* hybridization of the probe/primer derived from
10 the porcine retroviral sequence to the nucleic acid derived from donor, e.g., miniature swine, tissue. Alternatively, direct PCR analysis, using primers specific for porcine retroviral genes (e.g., genes comprising the nucleotide sequence shown in SEQ ID NO: 1, 2, or 3), can be used to detect the presence or absence of the genetic lesion in the porcine retroviral genome.

Miniature swine retroviral sequences of the invention can also be use to detect viral
15 recombinants within the genome, or in the circulation, cells, or transplanted tissue, between the porcine retrovirus and other endogenous human viruses or opportunistic pathogens (e.g. cytomegalovirus) of the immunocompromised transplant recipient. For example, pieces of the viral genome can be detected via PCR or via hybridization, e.g., Southern or Northern blot hybridization, using sequences derived from SEQ ID NO: 1, 2, or 3 as primers for
20 amplification or probes for hybridization.

Miniature swine retroviral sequences of the invention, e.g., PCR primers, allow quantitation of activated virus. Sequences of the invention also allow histologic localization (e.g., by *in situ* hybridization) of activated retrovirus. Localization allows clinicians to determine whether a graft should be removed as a source of potential retroviral infection of
25 the human host or whether the retroviral infection was localized outside the graft.

Sequences of the invention, e.g., PCR primers, allow the detection of actively replicating virus, e.g., by using reverse transcribed PCR techniques known in the art. Standard techniques for reverse transcriptase measurements are often complicated, species-specific, and are of low sensitivity and specificity, and false positive results may develop
30 using full-length probes for Southern and Northern molecular blotting. Sequences of the invention allow for sensitive and specific assays for the activation of virus and this will allow performance of a wide variety of tests, some of which are outlined below.

The invention provides for the testing and development of donor animals having reduced numbers of intact proviral insertions. It also provides for the testing of
35 immunosuppressive regimens less likely to provide the conditions for active replication of retrovirus. Conditions likely to activate one retrovirus are generally more likely to activate other viruses including unknown retroviruses and known human pathogens including cytomegalovirus, hepatitis B and C viruses, Human Immunodeficiency Viruses (I and II). Given the availability of preventative therapies for these infections, these therapies could be

used prophylactically in patients known to be susceptible to the activation of porcine retrovirus.

5 The miniature swine retroviral sequences of the invention can be used to measure the response of the miniature swine retroviral infection in humans to therapy, e.g., immunomodulatory or antiviral therapy, e.g., antiviral agents, e.g., antiviral antibiotics. With HIV, susceptibility to antiviral antibiotics is determined by the genetic sequence of the reverse transcriptase gene (RT pol region) and other genes. The ability to determine the exact sequence of the retroviral genes will allow the detection of mutations occurring during infection which would then confer resistance of this virus to antiviral agents. Primers, e.g.,
10 for the RT-pol region, of the invention can be used to detect and to sequence clinical viral isolates from patients which have developed mutations by PDQ method described herein. The primers of the invention can also be used to determine whether tumor cells, e.g., cancer cells, e.g. lymphoma or hepatocellular carcinoma, developing in xenograft recipients contain porcine retroviral elements.

15 The porcine retroviral sequences of the invention can also be used to detect other homologous retroviruses and to determine whether these are the same or different as compared to the Tusukuba-1 retroviral sequences. For example, within a species, the polymerase genes are highly conserved. PCR assays aimed at the gag-pol region followed by sequence analysis allow for this detection of homologous viruses. The appropriate regions of the Tsukuba-1 virus can be determined by using sequences derived from SEQ ID NO:1,
20 described herein, to identify additional 5' and 3' viral genomic sequences. As is discussed elsewhere herein, the sequences from SEQ ID NO: 1 were used to obtain the sequence of the PK-15 retroviral insert (SEQ ID NO:2) and of a retroviral insertion in a miniature swine (SEQ ID NO:3).

25 Miniature swine retroviral sequences of the invention can be used to screen donor animals and xenograft recipients after transplantation both for infection, and as a measure of the appropriate level of immune suppression, regarding susceptibility to infection. Physicians, medical staff, family, or individuals who come into contact with graft recipients, and others, can be screened for infection with virus derived from the xenograft recipient.
30 Members of the population in general can also be screened. Such screening can be used for broad epidemiologic studies of the community. These methods can help in meeting the requirements of the F.D.A. regarding enhancing the safety of the recipients and of the community to exposure to new viruses introduced into the community by xenograft transplantation.

35 As is shown in Suzuka et al., 1986, FEBS 198:339, the swine retroviruses such as the Tsukuba-1 genome can exist as a circular molecule. Upon cloning the circular molecule is generally cleaved to yield a linear molecule. As will be understood by one skilled in the art, the start point and end point of the resulting linear molecule, and the relative subregions of the viral sequence will of course vary with the point of cleavage. For example, in the Suzuka

et al. reference the LTR is shown to be in an internal fragment. This is indicated herein in that the order of gag, pol, env in SEQ ID NO 1 is shown as env, gag, pol, while elsewhere herein the order of these regions is given as the naturally occurring gag, pol, env order.

5 Primers Derived from the Porcine Retroviral (Tsukuba-1) Genome Sequence

A number of different primers useful in the methods of the invention have been described herein. One skilled in the art can identify additional primers from the viral sequence of SEQ ID NO:1 by using methods known in the art. For example, when trying to identify potentially useful primers one skilled in the art would look for sequences (sequences
10 should be between about 15 and 30 nucleotides in length) which hybridize to SEQ ID NO:1 with high melting temperature; have a balanced distribution of nucleotides, e.g., a balanced distribution of A, T, C and Gs; have a terminal C or G; do not self-hybridize or internally complement.

15 Use of Primers Derived from the Porcine Retroviral (Tsukuba-1) Genome Sequence

I. Testing of organs or cells prior to transplantation

Potential donor animals can be screened for active retroviral replication prior to being used in transplantation. This allows avoidance of animals undergoing active viral replication. Replicating virus is often infectious in 100% of recipients, while nonreplicating, latent
20 provirus generally causes infection in 5 to 25% of recipients.

II. Testing of recipients

Serial samples, e.g., of white blood cells, can be obtained from a graft recipient monthly, e.g., for the first month and every three months thereafter. Tissue biopsies obtained
25 for evaluation of graft function can be used to evaluate the activation of retroviral sequences or of the expression retroviral sequences in graft tissue. Samples can be screened for the presence of retrovirus infection both specifically for the homologous virus, for viral recombinants containing portions of the viral genome, and for other retroviruses, using, e.g., PCR primers for the pol region of the virus, which is the region most likely to be conserved.
30 If virus is detected, quantitative PCR can be used to determine the relative stability of viral production. Cells isolated from xenograft recipients can be tested by cocultivation with permissive human and porcine (e.g., pig fallopian tube, pig macrophage, or pig testis) cell lines known to contain endogenous viruses. Isolated virus will be tested for homology with the parental strain and for mutations which might affect susceptibility to antiviral agents, e.g.,
35 antiviral antibiotics.

III. Testing of surgical and medical personnel and family members of graft recipient

Samples, e.g., white blood cells, can be banked (archived) from the surgical and medical personnel and from family members of the recipient prior to transplantation and at

three months intervals for the first year and at least annually thereafter. Epidemiologic studies can be performed on these samples as well. These samples can be tested if the recipient becomes viremic or if unusual clinical manifestations are noted in these individuals.

5 IV. Testing of tumor cells

Tumor cells which develop from a graft, or a graft recipient, can be tested for the presence of active retrovirus and for proviruses.

V. Testing of patients

10 Patients can be retested for any significant change in clinical condition or for increased immune suppression of graft rejection which may be associated with an increased risk of viral activation.

Sequencing of the porcine retroviral (Tsukuba-1) genome

15 A clone (P λ 8.8) containing the 8060 bp XhoI porcine retrovirus (Tsukuba-1) insert was used to transfect competent *E. coli*, and DNA was isolated for sequencing. The strategy used to sequence the 8060 bp porcine retrovirus genome included a combination of procedures which are outlined below.

Random fragments (1-3 kb) of the clone (P λ 8.8) were generated by sonication. The
20 fragments were blunt-ended and were subcloned into the EcoRV site of the pBluescript SK vector. Plasmid DNA was prepared using a modified alkaline lysis procedure. DNA sequencing was performed using DyeDeoxy termination reactions (ABI). Base specific fluorescent dyes were used as labels. Sequencing reactions were analyzed on 4.75% polyacrylamide gels by an ABI 373A-S or 373S automated sequencer. Subsequent data
25 analysis was performed on Sequencer™ 3.0 software. The following internal sequencing primers were synthesized:

	AP1	5'	GATGAACAGGCAGACATCTG	3'	(SEQ ID NO:48)
	AP2	5'	CGCTTACAGACAAGCTGTGA	3'	(SEQ ID NO:49)
30	AP3	5'	AGAACAAAGGCTGGGAAAGC	3'	(SEQ ID NO:50)
	AP4	5'	ATAGGAGACAGCCTGAACTC	3'	(SEQ ID NO:51)
	AP5	5'	GGACCATTTGTCTGACCCTAT	3'	(SEQ ID NO:52)
	AP6	5'	GTCAACACCTATACCAGCTC	3'	(SEQ ID NO:53)
	AP7	5'	CATCTGAGGTATAGCAGGTC	3'	(SEQ ID NO:54)
35	AP8	5'	GCAGGTGTAGGAACAGGAAC	3'	(SEQ ID NO:55)
	AP9	5'	ACCTGTTGAACCATCCCTCA	3'	(SEQ ID NO:56)
	AP10	5'	CGAATGGAGAGATCCAGGTA	3'	(SEQ ID NO:57)
	AP11	5'	CCTGCATCACTTCTCTTACC	3'	(SEQ ID NO:58)
	AP12	5'	TTGCCTGCTTGTGGAATACG	3'	(SEQ ID NO:59)
40	AP13	5'	CAAGAGAAGAAGTGGGGAATG	3'	(SEQ ID NO:60)
	AP14	5'	CACAGTCGTACACCACGCAG	3'	(SEQ ID NO:61)
	AP15	5'	GGGAGACAGAAGAAGAAAGG	3'	(SEQ ID NO:62)
	AP16	5'	CGATAGTCATTAGTCCCAGG	3'	(SEQ ID NO:63)
	AP17	5'	TGCTGGTTTGCATCAAGACCG	3'	(SEQ ID NO:64)
45	AP18	5'	GTGCGAAAGGCATACCTGCT	3'	(SEQ ID NO:65)

	AP19	5'	ACAGAGCCTCTGCTAAGAAG	3'	(SEQ ID NO:66)
	AP20	5'	GCAGCTGTTGACAATCATC	3'	(SEQ ID NO:67)
	AP21	5'	TATGAGGAGAGGGCTTGA	3'	(SEQ ID NO:68)
	AP22	5'	AGCAGACGTGCTAGGAGGT	3'	(SEQ ID NO:69)
5	AP23	5'	TCCTCTTGCTGTTTGCATC	3'	(SEQ ID NO:70)
	AP24	5'	CAGACACTCAGAACAGAGAC	3'	(SEQ ID NO:71)
	AP25	5'	ACATCGTCTAACCCACCTAG	3'	(SEQ ID NO:72)
	AP26	5'	CTCGTTTCTGGTCATACCTGA	3'	(SEQ ID NO:73)
	AP27	5'	GAGTACATCTCTCTAGGCA	3'	(SEQ ID NO:74)
10	AP28	5'	TGCCTAGAGACATGTACTC	3'	(SEQ ID NO:4)
	AP29	5'	CCTCTTCTAGCCATTCCTTCA	3'	(SEQ ID NO:5)

The clone (Pλ8.8) containing the 8060 bp XhoI porcine retrovirus (Tsukuba-1) insert was deposited with ATCC on December 27, 1995 (ATCC Deposit No.97396).

15

Determination of the porcine retroviral (Tsukuba-1) copy number in a miniature swine

Total genomic DNA was isolated from miniature swine kidney by the methods known in the art. The isolated genomic DNA was digested with either EcoRI or HindIII restriction enzyme. The DNA digests were electrophoresed on an agarose gel, Southern blotted and
20 hybridized to the full-length, purified, Tsukuba-1 sequence (SEQ ID NO:1) under high stringency conditions (0.1 X SSC, 65°C). In both digested samples (EcoRI or HindIII) at least six copies of the high molecular fragments of the miniature swine genome (over 16 Kb in size) hybridized to SEQ ID NO:1, indicating the presence of homologous retroviral sequences in porcine DNA.

25

Susceptibility Testing by Polymerase Chain Reaction DNA Quantitation (PDQ)

Polymerase chain reaction (PCR) DNA quantitation (PDQ) susceptibility testing can be used to rapidly and directly measure nucleoside sensitivity of porcine retrovirus isolates. PCR can be used to quantitate the amount of porcine retroviral RNA synthesized after *in vitro*
30 infection of peripheral blood mononuclear cells. The relative amounts of porcine retroviral RNA in cell lysates from cultures maintained at different drug concentrations reflect drug inhibition of virus replication. With the PDQ method both infectivity titration and susceptibility testing can be performed on supernatants from primary cultures of peripheral blood mononuclear cells.

35

The PDQ experiments can be performed essentially as described by Eron et al., *PNAS USA* 89:3241-3245, 1992. Briefly, aliquots (150µl) of serial dilutions of virus sample can be used to infect 2×10^6 PHA-stimulated donor PBMCs in 1.5 ml of growth medium per well of a flat-bottom 24-well plate (Corning). Separate cell samples can be counted, harvested, and lysed at 48, 72 and 96 hr. Quantitative PCR and porcine retrovirus copy-number
40 determination can then be performed in duplicate on each lysate.

The results of a PDQ infectivity titration assay can be used to determine the virus dilution and length of culture time employed in a subsequent PDQ susceptibility test. These parameters should be chosen so that the yield of porcine retrovirus specific PCR product for

the untreated control infection would fall on the porcine retrovirus copy-number standard curve before the curve approached its asymptotic maximum, or plateau. PHA-stimulated donor PBMCs can be incubated with drug for 4 hr prior to infection. Duplicate wells in a 24-well plate should receive identical porcine retrovirus inocula for each drug concentration tested and for the untreated infected controls. Uninfected controls and drug toxicity controls should be included in each experiment. All cultures can be harvested and cells lysed for PCT after either 48 or 72 hr. Previously characterized isolates can be used as assay standards in each experiment.

Cell pellets can be lysed in various volumes of lysis buffer (50 mM KCl/10mM Tris•HCl, pH 8.3/2.5 mM MgCl₂/0.5% Nonidet P-40/0.5% Tween 20/0.01% proteinase K) to yield a concentration of 1.2×10^4 cell equivalents/ μ l. Uniformity to cell lysate DNA concentrations should be confirmed in representative experiments by enhancement of Hoechst 33258 fluorescence (Mini-Fluorometer, Hoefer).

A conserved primer pair can be synthesized according to the pol gene sequences. The primers can then be used to amplify a 1580-base pair fragment of the porcine retrovirus pol gene from 1.2×10^5 cell equivalents of lysate by using PCR (GeneAmp, Cetus) under standard conditions. Amplifications should be repeated if porcine retrovirus DNA is amplifiable from reagent controls.

Porcine retrovirus pol gene amplification products can be specifically detected and quantitated as described (Conway, B.C. (1990) in Techniques in HIV Research, (Aldovani & Walker, eds.) (Stockton, New York) pp.40-46). Heat-denatured PCR products can be hybridized in a Streptavidin-coated microtiter plate well with both biotinylated capture probe and horseradish peroxidase (HRP)-labeled detector probe [enzyme-linked oligonucleotide solution sandwich hybridization assay ((ELOSAs), DuPont Medical Products, Billerica, MA) for 60 min at 37°C. After extensive washing to remove all reactants except probe-DNA hybrids, an HRP chromogen, tetramethylbenzidine (TMBBlue, Transgenic Sciences, Worcester, MA), should be added to each well. The HRP-catalyzed color development should be stopped after 1 hr by addition of sulfuric acid to 0.65 M. Absorbance (OD) at 450 nm can be measured in an automated microtiter plate reader (SLT Labinstruments, Hillsborough, NC).

A standard curve of porcine retrovirus DNA copy number can be generated in each PCR by using a dilution series of cells containing one porcine proviral genome per cell.

Preparation of a miniature swine having a knockout of Tsukuba-1 viral sequence using isogenic DNA targeting vectors

Isogenic DNA, or DNA that is substantially identical in sequence between the targeting vector and the target DNA in the chromosomes, greatly increases the frequency for homologous recombination events and gene targeting efficiency. Using isogenic-DNA targeting vectors, targeting frequencies of 80% or higher can be achieved in mouse

embryonic stem cells. This is in contrast to non-isogenic DNA vectors which normally yield targeting frequencies of around 0.5% to 5%, i.e., approximately two orders of magnitude lower than isogenic DNA vectors. Isogenic DNA constructs are predominantly integrated into chromosomes by homologous recombination rather than random integration. As a consequence, targeted mutagenesis of viral sequences, e.g., viral genes, can be carried out in biological systems including zygotes, which do not lend themselves to the use of elaborate selection protocols, resulting in production of animals, e.g., miniature swine, free of, or having a reduced number of, activatable viral sequences. In order for the isogenic DNA approach to be feasible, targeting vectors should be constructed from a source of DNA that is identical to the DNA of the organism to be targeted. Ideally, isogenic DNA targeting is carried out in inbred strains of animals, e.g., inbred miniature swine, in which all genetic loci are homozygous. Any animal of that strain can serve as a source for generating isogenic targeting vectors. This protocol for isogenic gene targeting is outlined in TeRiele et al., PNAS 89:5128-5132, 1992 and PCT/US92/07184, herein incorporated by reference. A protocol for producing Tsukuba-1 knockout miniature swine is described briefly below.

An insertion vector is designed as described by Hasty and Bradley (Gene Targeting Vectors for Mammalian Cells, in Gene Targeting: A Practical Approach, ed, Alexandra L. Joyner, IRL Press 1993). Insertion vectors require that only one crossover event occur for integration by homologous recombination into the native locus. The double strand breaks, the two ends of the vector which are known to be highly recombinogenic, are located on adjacent sequences on the chromosome. The targeting frequencies of such constructions will be in the range of 30 to 50%. One disadvantage of insertion vectors, in general, concerns the sequence duplications that are introduced and that potentially make the locus unstable. All these constructions are made using standard cloning procedures.

Replacement vectors have also been extensively described by Hasty and Bradley. Conceptually more straight forward than the insertion vector, replacement vectors use an essentially co-linear fragment of a stretch of Tsukuba-1 genomic sequence. Preferably, the DNA sequence from which an isogenic replacement vector is constructed includes approximately 6 to 10 kb of uninterrupted DNA. Two crossovers, one on either side of the selectable marker causes the mutant targeting vector to become integrated and replace the wild-type gene.

Microinjection of the isogenic transgene DNA into one of the pronuclei of a porcine embryo at the zygote stage (one-cell embryo) is accomplished by modification of a protocol described earlier (Hammer et al. 1985, Nature 315, 680; Pursel et al. 1989, Science 244, 1281). The age and the weight of the donor pigs, e.g., haplotype specific mini-swine, are critical to success. Optimally, the animals are of age 8 to 10 months and weigh 70 to 85 lbs. This increases the probability of obtaining an adequate supply of one-cell embryos for microinjection of the transgenes. In order to allow for accurate timing of the embryo collections at this stage from a number of embryo donors, the gilts are synchronized using a

preparation of synthetic progesterone (Regumate). Hormone implants are applied to designated gilts 30 days prior to the date of embryo collection. Twenty days later, ten days prior to the date of collection, the implants are removed and the animals are treated with additional hormones to induce superovulation to increase the number of embryos for microinjection. Three days following implant removal, the animals are treated with 400 to 1000 IU of pregnant mare serum gonadotropin (PMSG) and with 750 IU of human chorionic gonadotropin (hCG) three to four days later. These animals are bred by artificial insemination (AI) on two consecutive days following injection of hCG.

Embryo collections are performed as follows: three days following the initial injection of hCG, the animals are anesthetized with an intramuscular injection of Telazol (3 mg/lb), Rompum (2 mg/lb) and Atropine (1 mg/lb). A midline laparotomy is performed and the reproductive tract exteriorized. Collection of the zygotes is performed by cannulating the ampulla of the oviduct and flushing the oviduct with 10 to 15 ml phosphate buffered saline, prewarmed to 39° C. Following the collection the donor animals are prepared for recovery from surgery according to USDA guidelines. Animals used twice for embryo collections are euthanized according to USDA guidelines.

Injection of the transgene DNA into the pronuclei of the zygotes is carried out as summarized below: Zygotes are maintained in medium HAM F-12 supplemented with 10% fetal calf serum at 38° C in 5% CO₂ atmosphere. For injection the zygotes are placed into BMOC-2 medium, centrifuged at 13,000 g to partition the embryonic lipids and visualize the pronuclei. The embryos are placed in an injection chamber (depression slide) containing the same medium overlaid with light paraffin oil. Microinjection is performed on a Nikon Diaphot inverted-microscope equipped with Nomarski optics and Narishige micromanipulators. Using 40x lens power the embryos are held in place with a holding pipette and injected with a glass needle which is back-filled with the solution of DNA containing the transgenic element, e.g., a mutant viral gene (2 µg/ml). Injection of approximately 2 picoliters of the solution (4 femptograms of DNA), which is equivalent to around 500 copies of the transgenic element, e.g., a mutant viral gene, is monitored by the swelling of the pronucleus by about 50%. Embryos that are injected are placed into the incubator prior to transfer to recipient animals.

Recipient animals are prepared similarly to the donor animals, but not superovulated. Prior to the transfer of the injected embryos, recipient gilts are anesthetized, the abdomen opened surgically by applying a longitudinal incision and the ovaries exteriorized. The oviduct ipsilateral to the ovary with the larger number of corpus lutei is flushed, the embryos checked to evaluate if the animals is reproductively sound. Approximately 4 to 6 zygotes injected with the transgenic element, e.g., a mutant viral gene, are transferred to the flushed oviduct, the abdominal incision sutured and the animals placed in a warm area for recovery. The status of the pregnancy is monitored by ultrasound starting at day 25, or approximately

one week following the expected date of implantation. Pregnant recipients are housed separately until they are due to farrow.

Newborn piglets are analyzed for integration of the transgenic element into chromosomal DNA. Genomic DNA is extracted from an ear punch or a blood sample and initial screening is performed using PCR. Animals that are potentially transgenic element-positive are confirmed by Southern analysis. Transgenic founder animals are subjected to further analysis regarding the locus of transgenic element integration using Southern analysis.

The isolation and sequencing of an endogenous swine retroviral insert and of a retroviral insert in porcine PK-15 cells

Cloning of PK15 and PAL endogenous retroviruses

I. Poly A⁺ RNA isolation

Peripheral blood lymphocytes (PBLs) were prepared from haplotype d/d miniswine using standard protocols known in the art. The PBLs were cultured in the presence of 1% phytohemagglutinin (PHA) for about 84 hours. The activated PBLs were collected and total RNA was isolated using commercially available kits, such as Gentra's (Minneapolis, Minnesota) PUREscript Kit. Poly A⁺RNA was isolated from the total RNA using another commercially available product, Dynal Dynabeads (Lake Success, NY). Northern analysis of the RNA using a pig retroviral probe confirmed the presence of potentially full-length retroviral genome RNA. RNA from PK15 cells was isolated using similar protocols.

II. Construction of the cDNA libraries

Using Superscript Choice System (Life Technologies Ltd, Gibco BRL, Gaithersburg, MD) for cDNA Synthesis, a cDNA library was constructed using oligo dT to make the first strand cDNA. The use of Superscript reverse transcriptase was important in order to obtain full-length retroviral (RV) cDNAs, due to the length of the RV RNA. The cDNA library was enriched for large cDNA fragments by size selecting >4 kb fragments by gel electrophoresis. The cDNAs were cloned into Lambda ZAP Express (Clontech Laboratories, Inc. Palo Alto, CA), which is one of the few commercially available cDNA vectors that would accept inserts in the 1-12kb range.

III. Screening of the cDNA libraries

0.75 - 1.2 x 10⁶ independent clones were screened using either gag and pol or gag and env probes. Double positive clones were further purified until single isolates were obtained (1 or 2 additional rounds of screening).

IV. Characterization of the clones

Between 18 and 30 double positive clones were selected for evaluation. Lambda DNA was prepared using standard protocols, such as the Lambda DNA Kit (Qiagen Inc., Chatsworth, CA). The clones were analyzed by PCR to check for (a) RV genes, and (b) determine the size of insert and LTR regions. Restriction digests were also done to confirm the size of insert and to attempt to categorize the clones. Clones containing the longest inserts and having consistent and predicted PCR data were sequenced.

Development of a PCR-based assay for the detection of the presence of an endogenous retrovirus in cells, tissues, organs, miniswine or recipient hosts (e.g., primates, humans)

- Using a commercially available computer software program (such as RightPrimer, Oligo 4.0, MacVector or Geneworks), one can analyze sequences disclosed herein for the selection of PCR primer pairs. The criteria for the general selection of primer pairs includes:
- a. The T_m of each primer is between 65-70°C
 - b. The T_m 's for each pair differ by no more than 3°C
 - c. The PCR fragment is between 200-800 bp in length
 - d. There are no repeats, self complementary bases, primer-dimer issues, etc for each pair
- A. Additional criteria for: A pig-specific PCR assay
- a. Primers are selected within porcine-specific regions of the sequence -- such as within gag, env, or U3. Porcine-specific primers are defined as sequences which overall have <70% homology to the corresponding region in human, mouse and primate retroviruses. In addition, the last five bases at the 3' end of the primer should be unique to the pig retroviral sequence.
 - b. Primers should have no more than one or two mismatched bases based on the miniswine, and retroviral sequences disclosed herein. These mismatched bases should not be within the last three or four bases of the 3' end of the primer.
- B. Additional criteria for: Miniswine-specific PCR assay
- a. Primers are selected such that there are at least one or two mismatches between miniswine and domestic pig sequences. At least one of these mismatches should be located within the last three or four bases at the 3' end of the primer. Preferably, these mismatches would be a change from either a G or C in miniswine to either an A or T in domestic pig.

RT-PCR Strategy

There are a number of commercially available RT-PCR Kits for routine amplification of fragments. Several primer pairs should be tested to confirm T_m and specificity. Location of primers within the sequence depends in part on what question is being answered. RT-PCR

should answer questions about expression and presence of RV sequences. PCR will not necessarily answer the question of whether the retroviral sequence is full-length or encodes a replication competent retrovirus. A positive signal in these tests only says there is RV sequence present. Indication of the possibility of full-length viral genomes being present can be obtained by performing long PCR using primers in U5 and U3. A commercial kit for long RT-PCR amplification is available (Takara RNA LA PCR Kit). Confirmation of full-length viral genomes requires infectivity studies and/or isolation of viral particles.

Northern analyses would complement RT-PCR data. Detection of bands at the predicted size of full-length viral genomes with hybridization probes from env, U3 or U5 would provide stronger evidence. The presence of other small bands hybridizing would indicate the amount of defective viral fragments present.

ELISA-BASED ASSAY TO DETECT THE PRESENCE OF PORCINE RETROVIRAL PROTEINS, POLYPEPTIDES OR PEPTIDES

In addition to the use of nucleic acid-based, e.g., PCR-based assays, to detect the presence of retroviral sequences, ELISA based assays can detect the presence of porcine retroviral proteins, polypeptides and peptides.

The basic steps to developing an ELISA include (a) generation of porcine retroviral specific peptides, polypeptides and proteins; (b) generation of antibodies which are specific for the porcine retroviral sequences; (c) developing the assay.

Using the retroviral sequences disclosed herein, antigenic peptides can be designed using computer based programs such as MacVector or Geneworks to analyse the retroviral sequences. Alternatively, it is possible to express the porcine retroviral sequences in gene expression systems and to purify the expressed polypeptides or proteins. After synthesis, the peptides, polypeptides or proteins are used to immunize mice or rabbits and to develop serum containing antibodies.

Having obtained the porcine retroviral specific antibodies the ELISA can be developed as follows. ELISA plates are coated with a volume of polyclonal or monoclonal antibody (capture antibody) which is reactive with the analyte to be tested. Such analytes include porcine retroviruses or retroviral proteins such as env or p24. The ELISA plates are then incubated at 4°C overnight. The coated plates are then washed and blocked with a volume of a blocking reagent to reduce or prevent non-specific hybridization. Such blocking reagents include bovine serum albumin (BSA), fetal bovine serum (FBS), milk, or gelatin. The temperature for the blocking process is 37°C. Plates can be used immediately or stored frozen at -20°C until needed. The plates are then washed, loaded with a serial dilution of the analyte, incubated at 37°C, and washed again. Bound analyte is detected using a detecting antibody. Detecting antibodies include enzyme-linked, fluoresceinated, biotin-conjugated or other tagged polyclonal or monoclonal antibodies which are reactive with the analyte. If

monoclonal antibodies are used the detecting antibody should recognize an epitope which is different from the capture antibody.

Other Embodiments

5 In another aspect, the invention provides a substantially pure nucleic acid having, or comprising, a nucleotide sequence which encodes a swine or miniature swine, e.g., a Tsukuba-1 retroviral gag polypeptide.

10 In preferred embodiments: the nucleic acid is or includes the nucleotide sequence from nucleotides 2452-4839 of SEQ ID NO:1; the nucleic acid is at least 60%, 70%, 80%, 90%, 95%, 98%, or 99% homologous with a nucleic acid sequence corresponding to nucleotides 2452-4839 of SEQ ID NO:1; or by a sequence which, hybridizes under high stringency conditions to nucleotides 2452-4839 of SEQ ID NO:1; the nucleic acid includes a fragment of SEQ ID NO:1 which is at least 25, 50, 100, 200, 300, 400, 500, or 1,000 bases in length; the nucleic acid differs from the nucleotide sequence corresponding to nucleotides 2452-4839 of SEQ ID NO:1 due to degeneracy in the genetic code; the nucleic acid differs
15 from the nucleic acid sequence corresponding to nucleotides 2452-4839 of SEQ ID NO:1 by at least one nucleotide but by less than 5, 10, 15 or 20 nucleotides and preferably which encodes an active peptide.

20 In yet another preferred embodiment, the nucleic acid of the invention hybridizes under stringent conditions to a nucleic acid probe corresponding to at least 12 consecutive nucleotides from nucleotides 2452-4839 of SEQ ID NO:1, or more preferably to at least 20 consecutive nucleotides from nucleotides 2452-4839 of SEQ ID NO:1, or more preferably to at least 40 consecutive nucleotides from nucleotides 2452-4839 of SEQ ID NO:1.

25 In another aspect, the invention features, a purified recombinant nucleic acid having at least 50%, 60%, 70%, 80%, 90%, 95%, 98%, or 99% homology with a nucleotide sequence corresponding to nucleotides 2452-4839 of SEQ ID NO:1.

30 The invention also provides a probe or primer which includes or comprises a substantially purified oligonucleotide. The oligonucleotide includes a region of nucleotide sequence which hybridizes under stringent conditions to at least 10 consecutive nucleotides of sense or antisense sequence from nucleotides 2452-4839 of SEQ ID NO:1, or naturally occurring mutants thereof. In preferred embodiments, the probe or primer further includes a label attached thereto. The label can be, e.g., a radioisotope, a fluorescent compound, an enzyme, and/or an enzyme co-factor. Preferably the oligonucleotide is at least 10 and less than 20, 30, 50, 100, or 150 nucleotides in length. Preferred primers of the invention include oligonucleotides having a nucleotide sequence shown in any of SEQ ID NOs:32-37.

35 The invention involves nucleic acids, e.g., RNA or DNA, encoding a polypeptide of the invention. This includes double stranded nucleic acids as well as coding and antisense single strands.

In another aspect, the invention provides a substantially pure nucleic acid having, or comprising, a nucleotide sequence which encodes a swine or miniature swine, e.g., a Tsukuba-1 retroviral pol polypeptide.

In preferred embodiments: the nucleic acid is or includes the nucleotide sequence
5 corresponding to nucleotides 4871-8060 of SEQ ID NO:1; the nucleic acid is at least 60%,
70%, 80%, 90%, 95%, 98%, or 99% homologous with a nucleic acid sequence corresponding
to nucleotides 4871-8060 of SEQ ID NO:1; or by a sequence which, hybridizes under high
stringency conditions to nucleotides 4871-8060 of SEQ ID NO:1; the nucleic acid includes a
fragment of SEQ ID NO:1 which is at least 25, 50, 100, 200, 300, 400, 500, or 1,000 bases in
10 length; the nucleic acid differs from the nucleotide sequence corresponding to nucleotides
4871-8060 of SEQ ID NO:1 due to degeneracy in the genetic code; the nucleic acid differs
from the nucleic acid sequence corresponding to nucleotides 4871-8060 of SEQ ID NO:1 by
at least one nucleotide but by less than 5, 10, 15 or 20 nucleotides and preferably which
encodes an active peptide.

15 In yet another preferred embodiment, the nucleic acid of the invention hybridizes
under stringent conditions to a nucleic acid probe corresponding to at least 12 consecutive
nucleotides from nucleotides 4871-8060 of SEQ ID NO:1, or more preferably to at least 20
consecutive nucleotides from nucleotides 4871-8060 of SEQ ID NO:1, or more preferably to
at least 40 consecutive nucleotides from nucleotides 4871-8060 of SEQ ID NO:1.

20 In another aspect, the invention features, a purified recombinant nucleic acid having at
least 50%, 60%, 70%, 80%, 90%, 95%, 98%, or 99% homology with a nucleotide sequence
corresponding to nucleotides 4871-8060 of SEQ ID NO:1.

The invention also provides a probe or primer which includes or comprises a
substantially purified oligonucleotide. The oligonucleotide includes a region of nucleotide
25 sequence which hybridizes under stringent conditions to at least 10 consecutive nucleotides
of sense or antisense sequence from nucleotides 4871-8060 of SEQ ID NO:1, or naturally
occurring mutants thereof. In preferred embodiments, the probe or primer further includes a
label attached thereto. The label can be, e.g., a radioisotope, a fluorescent compound, an
enzyme, and/or an enzyme co-factor. Preferably the oligonucleotide is at least 10 and less
30 than 20, 30, 50, 100, or 150 nucleotides in length. Preferred primers of the invention include
oligonucleotides having a nucleotide sequence shown in any of SEQ ID NOs:38-47.

The invention involves nucleic acids, e.g., RNA or DNA, encoding a polypeptide of
the invention. This includes double stranded nucleic acids as well as coding and antisense
single strands.

35 In another aspect, the invention provides a substantially pure nucleic acid having, or
comprising, a nucleotide sequence which encodes a swine or miniature swine, e.g., a
Tsukuba-1 retroviral env polypeptide.

In preferred embodiments: the nucleic acid is or includes the nucleotide sequence
corresponding to nucleotides 2-1999 of SEQ ID NO:1; the nucleic acid is at least 60%, 70%,

80%, 90%, 95%, 98%, or 99% homologous with a nucleic acid sequence corresponding to nucleotides 2-1999 of SEQ ID NO:1; or by a sequence which, hybridizes under high stringency conditions to nucleotides 2-1999 of SEQ ID NO:1; the nucleic acid includes a fragment of SEQ ID NO:1 which is at least 25, 50, 100, 200, 300, 400, 500, or 1,000 bases in length; the nucleic acid differs from the nucleotide sequence corresponding to nucleotides 2-1999 of SEQ ID NO:1 due to degeneracy in the genetic code; the nucleic acid differs from the nucleic acid sequence corresponding to nucleotides 2-1999 of SEQ ID NO:1 by at least one nucleotide but by less than 5, 10, 15 or 20 nucleotides and preferably which encodes an active peptide.

10 In yet another preferred embodiment, the nucleic acid of the invention hybridizes under stringent conditions to a nucleic acid probe corresponding to at least 12 consecutive nucleotides from nucleotides 2-1999 of SEQ ID NO:1, or more preferably to at least 20 consecutive nucleotides from nucleotides 2-1999 of SEQ ID NO:1, or more preferably to at least 40 consecutive nucleotides from nucleotides 2-1999 of SEQ ID NO:1.

15 In another aspect, the invention features, a purified recombinant nucleic acid having at least 50%, 60%, 70%, 80%, 90%, 95%, 98%, or 99% homology with a nucleotide sequence corresponding to nucleotides 2-1999 of SEQ ID NO:1.

The invention also provides a probe or primer which includes or comprises a substantially purified oligonucleotide. The oligonucleotide includes a region of nucleotide sequence which hybridizes under stringent conditions to at least 10 consecutive nucleotides of sense or antisense sequence from nucleotides 2-1999 of SEQ ID NO:1, or naturally occurring mutants thereof. In preferred embodiments, the probe or primer further includes a label attached thereto. The label can be, e.g., a radioisotope, a fluorescent compound, an enzyme, and/or an enzyme co-factor. Preferably the oligonucleotide is at least 10 and less than 20, 30, 50, 100, or 150 nucleotides in length. Preferred primers of the invention include oligonucleotides having a nucleotide sequence shown in any of SEQ ID NOs:6-31.

The invention includes nucleic acids, e.g., RNA or DNA, encoding a polypeptide of the invention. This includes double stranded nucleic acids as well as coding and antisense single strands.

30 Included in the invention are: allelic variations, natural mutants, induced mutants, that hybridize under high or low stringency conditions to the nucleic acid of SEQ ID NO:1, 2, or 3 (for definitions of high and low stringency see Current Protocols in Molecular Biology, John Wiley & Sons, New York, 1989, 6.3.1 - 6.3.6, hereby incorporated by reference).

35 The invention also includes purified preparations of swine or miniature swine retroviral polypeptides, e.g., gag pol, or env polypeptides, or fragments thereof, preferably biologically active fragments, or analogs, of such polypeptides. In preferred embodiments: the polypeptides are miniature swine retroviruses polypeptides; the polypeptides are Tsukuba polypeptides; the polypeptides are gag, pol, or env polypeptides encoded by SEQ ID NO:1

or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement, or naturally occurring variants thereof.

A biologically active fragment or analog is one having any in vivo or in vitro activity which is characteristic of the Tsukuba-1 polypeptides described herein, or of other naturally occurring Tsukuba-1 polypeptides. Fragments include those expressed in native or endogenous cells, e.g., as a result of post-translational processing, e.g., as the result of the removal of an amino-terminal signal sequence, as well as those made in expression systems, e.g., in CHO cells. A useful polypeptide fragment or polypeptide analog is one which exhibits a biological activity in any biological assay for Tsukuba-1 polypeptide activity. Most preferably the fragment or analog possesses 10%, preferably 40%, or at least 90% of the activity of Tsukuba-1 polypeptides, in any in vivo or in vitro Tsukuba-1 polypeptide assay.

In order to obtain a such polypeptides, polypeptide-encoding DNA can be introduced into an expression vector, the vector introduced into a cell suitable for expression of the desired protein, and the peptide recovered and purified, by prior art methods. Antibodies to the polypeptides can be made by immunizing an animal, e.g., a rabbit or mouse, and recovering antibodies by prior art methods.

The invention also features a purified nucleic acid, which has least 60%, 70%, 72%, more preferably at least 85%, more preferably at least 90%, more preferably at least 95%, most preferably at least 98%, 99% or 100% sequence identity or homology with SEQ ID NO:1 or its complement, SEQ ID NO: 2 or its complement, or SEQ ID NO: 3 or its complement.

In preferred embodiments the nucleic acid is other than the entire retroviral genome of SEQ ID NO:1 or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement, e.g., it is at least 1 nucleotide longer, or at least 1 nucleotide shorter, or differs in sequence at at least one position. E.g., the nucleic acid is a fragment of the sequence of SEQ ID NO:1 or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement, or it includes sequence additional to that of SEQ ID NO:1, or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement.

In preferred embodiments: the sequence of the nucleic acid differs from the corresponding sequence of SEQ ID NO: 1 or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement, by 1, 2, 3, 4, or 5 base pairs; the sequence of the nucleic acid differs from the corresponding sequence of SEQ ID NO: 1 or its complement, SEQ ID NO:2 or its complement, or SEQ ID NO:3 or its complement, by at least 1, 2, 3, 4, or 5 base pairs but less than 6, 7, 8, 9, or 10 base pairs.

In other preferred embodiments: the nucleic acid is at least 10, more preferably at least 15, more preferably at least 20, most preferably at least 25, 30, 50, 100, 1000, 2000, 4000, 6000, or 8060 nucleotides in length; the nucleic acid is less than 15, more preferably less than 20, most preferably less than 25, 30, 50, 100, 1000, 2000, 4000, 6000, or 8060 nucleotides in length.

Equivalents

Those skilled in the art will be able to recognize, or be able to ascertain using no more than routine experimentation, numerous equivalents to the specific procedures described
5 herein. Such equivalents are considered to be within the scope of this invention and are covered by the following claims.

SEQUENCE LISTING

- 5 (1) GENERAL INFORMATION:
- (i) APPLICANT: Jay A. Fishman
- (ii) TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
AND METHODS OF USE
- 10 (iii) NUMBER OF SEQUENCES: 74
- (iv) CORRESPONDENCE ADDRESS:
- 15 (A) ADDRESSEE: LAHIVE & COCKFIELD, LLP
(B) STREET: 60 State Street
(C) CITY: Boston
(D) STATE: Massachusetts
(E) COUNTRY: USA
(F) ZIP: 02109-1875
- 20 (v) COMPUTER READABLE FORM:
- (A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
25 (D) SOFTWARE: PatentIn Release #1.0, Version #1.25
- (vi) CURRENT APPLICATION DATA:
- (A) APPLICATION NUMBER:
30 (B) FILING DATE:
- (vii) PRIOR APPLICATION DATA:
- (A) APPLICATION NUMBER: US 08/572,645
(B) FILING DATE: 14-DEC-1995
- 35 (viii) ATTORNEY/AGENT INFORMATION:
- (A) NAME: Louis Myers
(B) REGISTRATION NUMBER: 35,965
(C) REFERENCE/DOCKET NUMBER: MGP-038CP
- 40 (ix) TELECOMMUNICATION INFORMATION:
- (A) TELEPHONE: (617)227-7400
(B) TELEFAX: (617)227-5941
- 45 (2) INFORMATION FOR SEQ ID NO:1:
- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 8060 base pairs
(B) TYPE: nucleic acid
50 (C) STRANDEDNESS: single
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- 55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

	GTCGAAGGAA	TCTCCACCTG	GATCCATGCA	TCCCACGTTA	AGCCGGCGCC	ACCTCCCGAT	120
	TCGGGGTGGG	AAGCCGAAAA	GACTGAAAAT	CCCCTTAAGC	TTCGCCTCCA	TCGCGTGGTT	180
5	CCTTACTCTG	TCAATAACCT	CTCAGACTAA	TGGTATGCGC	ATAGGAGACA	GCCTGAACTC	240
	CCATAAACCC	TTATCTCTCA	CCTGGTTAAT	TACTGACTCC	GGCACAGGTA	TTAATATCAA	300
10	CAACACTCAA	GGGGAGGCTC	CTTTAGGAAC	CTGGTGGCCT	GATCTATACG	TTTGCCTCAG	360
	ATCAGTTATT	CCTAGTCTGA	CCTCACCCCC	AGATATCCTC	CATGCTCAGC	GATTTTATGT	420
	TTGCCCAGGA	CCACCAAATA	ATGGAAAACA	TTGCGGAAAT	CCCAGAGATT	TCTTTTGTA	480
15	ACAATGGAAC	TGTGTAACCT	CTAATGATGG	ATATTGGAAA	TGGCCAACCT	CTCAGCAGGA	540
	TAGGGTAAGT	TTTTCTTATG	TCAACACCTA	TACCAGCTCT	GGACAATTTA	ATTACCTGAC	600
20	CTGGATTAGA	ACTGGAAGCC	CCAAGTGCTC	TCCTTCAGAC	CTAGATTACC	TAAAAATAAG	660
	TTTCACTGAG	AAAGGAAAAC	AAGAAAATAT	CCTAAAATGG	GTAAATGGTA	TGTCTTGGGG	720
	AATGGTATAT	TATGGAGGCT	CGGGTAAACA	ACCAGGCTCC	ATTCTAACTA	TTCGCCTCAA	780
25	AATAAACCAG	CTGGAGCCTC	CAATGGCTAT	AGGACCAAAT	ACGGTCTTGA	CGGGTCAAAG	840
	ACCCCCAACC	CAAGGACCAG	GACCATCCTC	TAACATAACT	TCTGGATCAG	ACCCCACTGA	900
30	GTCTAGCAGC	ACGACTAAAA	TGGGGGCAAA	ACTTTTTAGC	CTCATCCAGG	GAGCTTTTCA	960
	AGCTCTTAAC	TCCACGACTC	CAGAGGCTAC	CTCTTCTTGT	TGGCTATGCT	TAGCTTTGGG	1020
	CCCACCTTAC	TATGAAGGAA	TGGCTAGAAG	AGGGAAATTC	AATGTGACAA	AAGAACATAG	1080
35	AGACCAATGC	ACATGGGGAT	CCCAAATAA	GCTTACCCTT	ACTGAGGTTT	CTGGAAAAGG	1140
	CACCTGCATA	GGAAAGGTTC	CCCCATCCCA	CCAACACCTT	TGTAACCACA	CTGAAGCCTT	1200
40	TAATCAAACC	TCTGAAAGTC	AATATCTGGT	ACCTGGTTAT	GACAGGTGGT	GGGCATGTAA	1260
	TACTGGATTA	ACCCCTTGTG	TTTCCACCTT	GGTTTTTAAC	CAAACTAAAG	ATTTTTCAT	1320
	TATGGTCCAA	ATTGTTCCCC	GAGTGTATTA	CTATCCCGAA	AAAGCAATCC	TTGATGAATA	1380
45	TGACTACAGA	AATCATCGAC	AAAAGAGAGA	ACCCATATCT	CTGACACTTG	CTGTGATGCT	1440
	CGGACTTGGA	GTGGCAGCAG	GTGTAGGAAC	AGGAACAGCT	GCCCTGGTCA	CGGGACCACA	1500
50	GCAGCTAGAA	ACAGGACTTA	GTAACCTACA	TCGAATTGTA	ACAGAAGATC	TCCAAGCCCT	1560
	AGAAAAATCT	GTCAGTAACC	TGGAGGAATC	CCTAACCTCC	TTATCTGAAG	TAGTCCTACA	1620
	GAATAGAAGA	GGGTTAGATT	TATTATTTCT	AAAAGAAGGA	GGATTATGTG	TAGCCTTGAA	1680
55	GGAGGAATGC	TGTTTTTATG	TGGATCATTC	AGGGGCCATC	AGAGACTCCA	TGAACAACT	1740
	TAGAGAAAGG	TTGGAGAAGC	GTCGAAGGGA	AAAGGAAACT	ACTCAAGGGT	GGTTTGAGGG	1800

	ATGGTTCAAC	AGGTCTCCTT	GGTTGGCTAC	CCTACTTTCT	GCTTTAACAG	GACCCTTAAT	1860
	AGTCCTCCTC	CTGTTACTCA	CAGTTGGGCC	ATGTATTATT	AACAAGTTAA	TTGCCTTCAT	1920
5	TAGAGAACGA	ATAAGTGCAG	TCCAGATCAT	GGTACTTAGA	CAACAGTACC	AAAGCCCGTC	1980
	TAGCAGGGAA	GCTGGCCGCT	AGCTCTACCA	GTTCTAAGAT	TAGAACTATT	AACAAGAGAA	2040
	GAAGTGGGGA	ATGAAAGGAT	GAAAATACAA	CCTAAGCTAA	TGAGAAGCTT	AAAATTGTTC	2100
10	TGAATTCCAG	AGTTTGTTCC	TTATAGGTAA	AAGATTAGGT	TTTTTGCTGT	TTTAAATAT	2160
	GCGGAAGTAA	AATAGGCCCT	GAGTACATGT	CTCTAGGCAT	GAACTTCTT	GAACTATTT	2220
15	GAGATAACAA	GAAAAGGGAG	TTTCTAACTG	CTTGTTTAGC	TTCTGTAAAA	CTGGTTGCGC	2280
	CATAAAGATG	TTGAAATGTT	GATACACATA	TCTTGGTGAC	AACATGTCTC	CCCCACCCCG	2340
	AAACATGCGC	AAATGTGTAA	CTCTAAACA	ATTTAAATTA	ATTGGTCCAC	GAAGCGCGGG	2400
20	CTCTCGAAGT	TTTAAATTGA	CTGGTTTGTG	ATATTTTGAA	ATGATTGGTT	TGTAAAGCGC	2460
	GGGCTTTGCT	GTGAACCCCA	TAAAAGCTGT	CCCGACTCCA	CACTCGGGGC	CGCAGTCCTC	2520
25	TACCCCTGCG	TGGTGTACGA	CTGTGGGCC	CAGCGCGCTT	GGAATAAAAA	TCCTCTTGCT	2580
	GTTTGCATCA	AGACCGCTTC	TCTGTAGTGA	TTAAGGGGAG	TCGCCTTTTC	CGAGCCTGGA	2640
	GGTTCTTTTT	GCTGGTCTTA	CATTTGGGGG	CTCGTCCGGG	ATCTGTCGCG	GCCACCCCTA	2700
30	ACACCCGAGA	ACCGACTTGG	AGGTAAAAAG	GATCCTCTTT	TTAACGTGTA	TGCATGTACC	2760
	GGCCGGCGTC	TCTGTTCTGA	GTGTCTGTTT	TCAGTGGTGC	GCGCTTTCGG	TTTGCAGCTG	2820
35	TCCTCTCAGG	CCGTAAGGGC	TGGGGGACTG	TGATCAGCAG	ACGTGCTAGG	AGGATCACAG	2880
	GCTGCTGCCC	TGGGGGACGC	CCCGGGAGGT	GAGGAGAGCC	AGGGACGCCT	GGTGGTCTCC	2940
	TACTGTCCGT	CAGAGGACCG	AATTCTGTTG	CTGAAGCGAA	AGCTTCCCCC	TCCGCGACCG	3000
40	TCCGACTCTT	TTGCCTGCTT	GTGGAATACG	TGGACGGGTC	ACGTGTGTCT	GGATCTGTTG	3060
	GTTTCTGTTT	TGTGTGTCTT	TGTCTTGTGT	GTCCTTGTCT	ACAGTTTTAA	TATGGGACAG	3120
45	ACGGTGACGA	CCCCTCTTAG	TTTGA CTCTC	GACCATTGGA	CTGAAGTTAA	ATCCAGGGCT	3180
	CATAATTTGT	CAGTTCAGGT	TAAGAAGGGA	CCTTGGCAGA	CTTTCTGTGT	CTCTGAATGG	3240
	CCGACATTCG	ATGTTGGATG	GCCATCAGAG	GGGACCTTTA	ATTCTGAGAT	TATCCTGGCT	3300
50	GTAAAGCAA	TTATTTTTCA	GACTGGACCC	GGCTCTCATC	CCGATCAGGA	GCCCTATATC	3360
	CTTACGTGGC	AAGATTTGGC	AGAGGATCCT	CCGCCATGGG	TTAAACCATG	GCTGAATAAG	3420
55	CCAAGAAAGC	CAGGTCCCCG	AATTCTGGCT	CTTGGAGAGA	AAAACAAACA	CTCGGCTGAA	3480
	AAAGTCAAGC	CCTCTCCTCA	TATCTACCCC	GAGATTGAGG	AACCACCGGC	TTGGCCGGAA	3540
	CCCCAATCTG	TTCCCCCACC	CCCTTATCTG	GCACAGGGTG	CCGCGAGGGG	ACCCTTTGCC	3600

	CCTCCTGGAG	CTCCGGCGGT	GGAGGGACCT	TCTGCAGGGA	CTCGGAGCCG	GAGGGGCGCC	3660
	ACCCCGGAGC	GGACAGACGA	GATCGCGACA	TTACCGCTGC	GCACGTACGG	CCCTCCCACA	3720
5	CCGGGGGGCC	AATTGCAGCC	CCTCCAGTAT	TGGCCCTTTT	CTTCTGCAGA	TCTCTATAAT	3780
	TGAAAACTA	ACCATCCCCC	TTTCTCGGAG	GATCCCCAAC	GCCTCACGGG	GTTGGTGGAG	3840
10	TCCCTTATGT	TCTCTCACCA	GCCTACTTGG	GATGATTGTC	AACAGCTGCT	GCAGACACTC	3900
	TTCACAACCG	AGGAGCGAGA	GAGAATTCTA	TTAGAGGCTA	GAAAAAATGT	TCCTGGGGCC	3960
	GACGGGCGAC	CCACGCGGTT	GCAAAATGAG	ATTGACATGG	GATTTCCCTT	AACTCGCCCC	4020
15	GGTTGGGACT	ACAACACGGC	TGAAGGTAGG	GAGAGCTTGA	AAATCTATCG	CCAGGCTCTG	4080
	GTGGCGGGTC	TCCGGGGCGC	CTCAAGACGG	CCCCTAATT	TGGCTAAGGT	AAGAGAAGTG	4140
20	ATGCAGGGAC	CGAATGAACC	CCCCTCTGTT	TTTCTTGAGA	GGCTCTTGGA	AGCCTTCAGG	4200
	CGGTACACCC	CTTTTGATCC	CACCTCAGAG	GCCCCAAAAG	CCTCAGTGGC	TTTGGCCTTT	4260
	ATAGGACAGT	CAGCCTTGGA	TATTAGAAAG	AAGCTTCAGA	GA CTGGAAGG	GTTACAGGAG	4320
25	GCTGAGTTAC	GTGATCTAGT	GAAGGAGGCA	GAGAAAGTAT	ATTACAAAAG	GGAGACAGAA	4380
	GAAGAAAGGG	AACAAAGAAA	AGAGAGAGAA	AGAGAGGAAA	GGGAGGAAAG	ACGTAATAAA	4440
30	CGGCAAGAGA	AGAATTTGAC	TAAGATCTTG	GCTGCAGTGG	TTGAAGGGAA	AAGCAATACG	4500
	GAAAGAGAGA	GAGATTTTAG	GAAAATTAGG	TCAGGCCCTA	GACAGTCAGG	GAACCTGGGC	4560
	AATAGGACCC	CACTCGACAA	GGACCAATGT	GCATATTGTA	AAGAAAGAGG	AACTGGGCA	4620
35	AGGAACTGCC	CCAAGAAGGG	AAACAAAGGA	CCAAGGATCC	TAGCTCTAGA	AGAAGATAAA	4680
	GATTAGGGGA	GACGGGGTTC	GGACCCCCCTC	CCCGAGCCCA	GGGTAAC TTT	GAAGGTGGAG	4740
40	GGGCAACCAG	TTGAGTTCCT	GGTTGATACC	GGAGCGAAAC	ATTCAGTGCT	ACTACAGCCA	4800
	TTAGGAAAAC	TAAAAGATAA	AAAATCCTGG	GTGATGGGTG	CACAGGGCAA	CAACAGTATC	4860
	CATGGACTAC	CCGAAGACAG	TTGACTTG GG	AGTGGGACGG	GTAACCCACT	CGTTTCTGGT	4920
45	CATACCTGAG	TGCCCAGCAC	CCCTCTTAGG	TAGAGACTTA	TTGACCAAGA	TGGGAGCACA	4980
	AATTTCTTTT	GAACAAGGGA	AACCAGAAGT	GTCTGCAAAT	AACAAACCTA	TCACTGTGTT	5040
50	GACCCTCCAA	TTAGATGACG	AATATCGACT	ATACTCTCCC	CTAGTAAAGC	CTGATCAAAA	5100
	TATACAATTC	TGGTTGGAAC	AGTTTCCCCA	AGCCTGGGCA	GAAACCGCAG	GGATGGGTTT	5160
	GGCAAAGCAA	GTTCCCCCAC	AAGTTATTCA	ACTGAAGGCC	AGTGCCACAC	CAGTGTCAGT	5220
55	CAGACAGTAC	CCCTTGAGTA	AAGAAGCTCA	AGAAGGAATT	CGGCCGCATG	TCCAAAGATT	5280
	AATCCAACAG	GGCATCCTAG	TTCCTGTCCA	ATCTCCCTGG	AATACTCCCC	TGCTACCGGT	5340

	TAGAAAGCCT	GGGACTAATG	ACTATCGACC	AGTACAGGAC	TTGAGAGAGG	TCAATAAACG	5400
	GGTGCAGGAT	ATACACCCAA	CAGTCCCGAA	CCCTTATAAC	CTCTTGTGTG	CTCTCCCACC	5460
5	CCAACGGAGC	TGGTATACAG	TATTGGACTT	AAAGGATGCC	TTCTTCTGCC	TGAGATTACA	5520
	CCCCACTAGC	CAACCACTTT	TTGCCTTCGA	ATGGAGAGAT	CCAGGTACGG	GAAGAACCGG	5580
	GCAGCTCACC	TGGACCCGAC	TGCCCCAAGG	GTTCAAGAAC	TCCCCGACCA	TCTTTGACGA	5640
10	AGCCCTACAC	AGAGACCTGG	CCAACTTCAG	GATCCAACAC	CCTCAGGTGA	CCCTCCTCCA	5700
	GTACGTGGAT	GACCTGCTTC	TGGCGGGAGC	CACCAAACAG	GACTGCTTAG	AAGGCACGAA	5760
15	GGCACTACTG	CTGGAATTGT	CTGACCTAGG	CTACAGAGCC	TCTGCTAAGA	AGGCCCAGAT	5820
	TTGCAGGAGA	GAGGTAACAT	ACTTGGGGTA	CAGTTTACGG	GACGGGCAGC	GATGGCTGAC	5880
	GGAGGCACGG	AAGAAAAC TG	TAGTCCAGAT	ACCGGCCCCA	ACCACAGCCA	AACAAATGAG	5940
20	AGAGTTTTTG	GGGACAGCTG	GATTTTGCAG	ACTGTGGATC	CCGGGGTTTG	CGACCTTAGC	6000
	AGCCCCACTC	TACCCGCTAA	CCAAAGAAAA	AGGGGAATTC	TCCTGGGCTC	CTGAGCACCA	6060
25	GAAGGCATTT	GATGCTATCA	AAAAGGCCCT	GCTGAGCGCA	CCTGCTCTGG	CCCTCCCTGA	6120
	CGTAACTAAA	CCCTTTACCC	TTTATGTGGA	TGAGCGTAAG	GGAGTAGCCC	GGGGAGTTTT	6180
	AACCCAAACC	CTAGGACCAT	GGAGAAGACC	TGTCGCCTAC	CTGTCAAAGA	AGCTCGATCC	6240
30	TGTAGCCAGT	GGTTGGCCCA	TATGCCTGAA	GGCTATCGCA	GCTGTGGCCA	TACTGGTCAA	6300
	GGACGCTGAC	AAATTGACTT	TGGGACAAGA	ATATAACTGT	AATAGCCCCC	CATGCATTGG	6360
35	AGAACATCGT	TCGGCAGCCC	CCAGACCGAT	GGATGACCAA	CGCCCGCATG	ACCCACTATC	6420
	AAAGCCTGCT	TCTCACAGAG	AGGGTCACGT	TCGCTCCACC	AACCGCTCTC	AACCCTGCCA	6480
	CTCTTCTGCC	TGAAGAGACT	GATGAACCAG	TGACTCATGA	TTGCCATCAA	CTATTGATTG	6540
40	AGGAGACTGG	GGTCCGCAAG	GACCTTACAG	ACATACCGCT	GACTGGAGAA	GTGCTAACCT	6600
	GGTTCACTGA	CGGAAGCAGC	TATGTGGTGG	AAGGTAAGAG	GATGGCTGGG	GCGGCGGTGG	6660
45	TGGACGGGAC	CCGCACGATC	TGGGCCAGCA	GCCTGCCGGG	AGGAACTTCA	GCACAAAAGG	6720
	CTGAGCTCAT	GGCCCTCACG	CAAGCTTTGC	GGCTGGCCGA	AGGGAAATCC	ATAAACATTT	6780
	ATACGGACAG	CAGGTATGCC	TTTGCGACTG	CACACGTACA	TGGGGCCATC	TATAAACAAA	6840
50	GGGGGTGCT	TACCTCAGCA	GGGAGGGAAA	TAAAGAACAA	AGAGGAAATT	CTAAGCCTAT	6900
	TAGAAGCCGT	ACATTTACCA	AAAAGGCTAG	CTATTATACA	CTGTCCTGGA	CATCAGAAAG	6960
55	CTAAAGATCT	CATATCCAGA	GGAAACCAGA	TGGCTGACCG	GGTTGCCAAG	CAGGCAGCCC	7020
	AGGGTGTTAA	CCTTCTGCCT	ATAATAGAAA	TGCCCAAAGC	CCCAGAACCC	AGACGACAGT	7080
	ACACCCTAGA	AGACTGGCAA	GAGATAAAAA	AGATAGACCA	TTCTCTGAGA	CTCCGGAAGG	7140

	GACCTGCTAT ACCTCAGATG GGAAGGAAAT CCTGCCCCAC AAAGAAGGGT TAGAATATGT	7200
	CCAACAAGAT ACATCGTCTA ACCCACCTAG GAACTAAACA CCTGCAGCAG TTGGTCAGAA	7260
5	CATCCCCCTTA TCATGTTCTG AGGCTACCAG GAGTGGCTGA CTCGGTGGTC AAACATTGTG	7320
	TGCCCTGCCA GCTGGTTAAT GCTAATCCTT CCAGAATGCC TCCAGGGAAG AGACTAAGGG	7380
10	GAAGCCACCC AGGCGCTCAC TGGGAAGTGG ACTTCACTGA GGTAAAGCCG GCTAAATATG	7440
	GAAACAAATA CCTATTGGTT TTTGTAGACA CCTTTTCAGG ATGGGTAGAG GCTTATCCTA	7500
	CTAAGAAAGA GACTTCAACC GTGGTAGCTA AAAAAATACT GGAAGAAATT TTTCCAAGAT	7560
15	TTGGAATACC TAAGGTAATA GGGTCAGACA ATGGTCCAGC TTTTGTGGCC CAGGTAAGTC	7620
	AGGGACTGGC CAAGATATTG GGGATTGATT GGAAACTGCA TTGTGCATAC AGACCCCAA	7680
20	GCTCAGGACA GGTAGAGAGG ATGAATAGAA CCATTAAAGA GACCCTTACT AAATTGACCG	7740
	CGGAGACTGG CGTTAATGAT TGGATAGCTC TCCTGCCCTT TGTGCTTTTT AGGGTTAGGA	7800
	ACACCCCTGG ACAGTTTGGG CTGACCCCTT ATGAATTACT CTACGGGGGA CCCCCCAT	7860
25	TGGTAGAAAT TGCTTCTGTA CATAGTGCTG ATGTGCTGCT TTCCCAGCCT TTGTTCTCTA	7920
	GGCTCAAGGC ACTTGAGTGG GTGAGACAAC GAGCGTGGAG GCAACTCCGG GAGGCCTACT	7980
30	CAGGAGGAGG AGACTTGCAG ATCCACATC GTTTCCAAGT GGGAGATTCA GTCTACGTTA	8040
	GACGCCACCG TGCAGGAAAC	8060

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7333 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

	CTACCCCTGC GTGGTGTACG ACTGTGGGCC CCAGCGCGCT TGGAATAAAA ATCCTCTTGC	60
50	TGTTTGCATC AAGACCGCTT CTTGTGAGTG ATTTGGGGTG TCGCCTCTTC CGAGCCCGGA	120
	CGAGGGGGAT TGTTCTTTTA CTGGCCTTTC ATTTGGTGCG TTGGCCGGGA AATCCTGCGA	180
55	CCACCCCTTA CACCCGAGAA CCGACTTGA GGTAAAGGGA TCCCCTTTGG AACATATGTG	240
	TGTGTGGGCC GGCCTCTCTG TTCTGAGTGT CTGTTTTCGG TGATGCGCGC TTTCGGTTTG	300
	CAGCTGTCCT CTCAGACCGT AAGGACTGGA GGAAGTGAT CAGCAGACGT GCTAGGAGGA	360

	TCACAGGCTG	CCACCCTGGG	GGACGCCCCG	GGAGGTGGGG	AGAGCCAGGG	ACGCTGGTG	420
	GTCTCCTACT	GTCCGTCAGA	GGACCGAGTT	CTGTTGTTGA	AGCGAAAGCT	TCCCCCTCCG	480
5	CGGCCGTCCG	ACTCTTTTGC	CTGCTTGTGG	AAGACGCGGA	CGGGTCGCGT	GTGTCTGGAT	540
	CTGTTGGTTT	CTGTTTCGTG	TGTCTTTGTC	TTGTGCGTCC	TTGTCTACAG	TTTTAATATG	600
10	GGACAGACAG	TGACTACCCC	CCTTAGTTTG	ACTCTCGACC	ATTGGACTGA	AGTTAGATCC	660
	AGGGCTCATA	ATTTGTCAGT	TCAGGTTAAG	AAGGGACCTT	GGCAGACTTT	CTGTGCCTCT	720
	GAATGGCCAA	CATTTCGATG	TGGATGGCCA	TCAGAGGGGA	CCTTTAATTC	TGAAATTATC	780
15	CTGGCTGTTA	AGGCAATCAT	TTTTCAGACT	GGACCCGGCT	CTCATCCTGA	TCAGGAGCCC	840
	TATATCCTTA	CGTGGCAAGA	TTTGGCAGAA	GATCCTCCGC	CATGGGTAA	ACCATGGCTA	900
20	AATAAACCAA	GAAAGCCAGG	TCCCCGAATC	CTGGCTCTTG	GAGAGAAAAA	CAAACACTCG	960
	GCCGAAAAAG	TCGAGCCCTC	TCCTCGTATC	TACCCCGAGA	TCGAGGAGCC	GCCGACTTGG	1020
	CCGGAACCCC	AACCTGTTCC	CCCACCCCT	TATCCAGCAC	AGGGTGCTGT	GAGGGGACCC	1080
25	TCTGCCCCTC	CTGGAGCTCC	GGTGGTGGAG	GGACCTGCTG	CCGGGACTCG	GAGCCGGAGA	1140
	GGCGCCACCC	CGGAGCGGAC	AGACGAGATC	GCGATATTAC	CGCTGCGCAC	CTATGGCCCT	1200
30	CCCATGCCAG	GGGGCCAATT	GCAGCCCCTC	CAGTATTGGC	CCTTTTCTTC	TGCAGATCTC	1260
	TATAATTGGA	AAACTAACCA	TCCCCCTTTC	TCGGAGGATC	CCCAACGCCT	CACGGGGTTG	1320
	GTGGAGTCCC	TTATGTTCTC	TCACCAGCCT	ACTTGGGATG	ATTGTCAACA	GCTGCTGCAG	1380
35	ACACTCTTCA	CAACCGAGGA	GCGAGAGAGA	ATTCTGTTAG	AGGCTAAAAA	AAATGTTCCCT	1440
	GGGGCCGACG	GGCGACCCAC	GCAGTTGCAA	AATGAGATTG	ACATGGGATT	TCCCTTGACT	1500
40	CGCCCCGGTT	GGGACTACAA	CACGGCTGAA	GGTAGGGAGA	GCTTGAAAAT	CTATCGCCAG	1560
	GCTCTGGTGG	CGGGTCTCCG	GGGCGCCTCA	AGACGGCCCA	CTAATTTGGC	TAAGGTAAGA	1620
	GAGGTGATGC	AGGGACCGAA	CGAACCTCCC	TCGGTATTTT	TTGAGAGGCT	CATGGAAGCC	1680
45	TTCAGGCGGT	TCACCCCTTT	TGATCCTACC	TCAGAGGCCC	AGAAAGCCTC	AGTGGCCCTG	1740
	GCCTTCATTG	GGCAGTCGGC	TCTGGATATC	AGGAAGAAAC	TTCAGAGACT	GGAAGGGTTA	1800
50	CAGGAGGCTG	AGTTACGTGA	TCTAGTGAGA	GAGGCAGAGA	AGGTGTATTA	CAGAAGGGAG	1860
	ACAGAAGAGG	AGAAGGAACA	GAGAAAAGAA	AAGGAGAGAG	AAGAAAGGGA	GGAAAGACGT	1920
	GATAGACGGC	AAGAGAAGAA	TTTGACTAAG	ATCTTGGCCG	CAGTGGTTGA	AGGGAAGAGC	1980
55	AGCAGGGAGA	GAGAGAGAGA	TTTTAGGAAA	ATTAGGTCAG	GCCCTAGACA	GTCAGGGAAC	2040
	CTGGGCAATA	GGACCCCACT	CGACAAGGAC	CAGTGTGCGT	ATTGTAAAGA	AAAAGGACAC	2100

	TGGGCAAGGA	ACTGCCCCAA	GAAGGGAAAC	AAAGGACCGA	AGGTCCTAGC	TCTAGAAGAA	2160
	GATAAAGATT	AGGGGAGACG	GGGTTCGGAC	CCCCTCCCCG	AGCCCAGGGT	AACTTTGAAG	2220
5	GTGGAGGGGC	AACCAGTTGA	GTTCTTGTT	GATACCGGAG	CGGAGCATTC	AGTGCTGCTA	2280
	CAACCATTAG	GAAAACATAA	AGAAAAAATA	TCCTGGGTGA	TGGGTGCCAC	AGGGCAACGG	2340
	CAGTATCCAT	GGACTACCCG	AAGAACCGTT	GACTTGGGAG	TGGGACGGGT	AACCCACTCG	2400
10	TTTCTGGTCA	TCCCTGAGTG	CCCAGTACCC	CTTCTAGGTA	GAGACTTACT	GACCAAGATG	2460
	GGAGCTCAAA	TTTCTTTTGA	ACAAGGAAGA	CCAGAAGTGT	CTGTGAATAA	CAAACCCATC	2520
15	ACTGTGTTGA	CCCTCCAATT	AGATGATGAA	TATCGACTAT	ATTCTCCCCA	AGTAAAGCCT	2580
	GATCAAGATA	TACAGTCCTG	GTTGGAGCAG	TTTCCCCAAG	CCTGGGCAGA	AACCGCAGGG	2640
	ATGGGTTTGG	CAAAGCAAGT	TCCCCACAG	GTTATTCAAC	TGAAGGCCAG	TGCTACACCA	2700
20	GATCAGTCA	GACAGTACCC	CTTGAGTAGA	GAGGCTCGAG	AAGGAATTG	GCCGCATGTT	2760
	CAAAGATTAA	TCCAACAGGG	CATCCTAGTT	CCTGTCCAAT	CCCCTTGGA	TACTCCCCTG	2820
25	CTACCGGTTA	GGAAGCCTGG	GACCAATGAT	TATCGACCAG	TACAGGACTT	GAGAGAGGTC	2880
	AATAAAAGGG	TGCAGGACAT	ACACCCAACG	GTCCCGAACC	CTTATAACCT	CTTGAGCGCC	2940
	CTCCCGCCTG	AACGGAAGT	GTACACAGTA	TTGGACTTAA	AAGATGCCTT	CTTCTGCCTG	3000
30	AGATTACACC	CCACTAGCCA	ACCACTTTTT	ACCTTCGAAT	GGAGAGATCC	AGGTACGGGA	3060
	AGAACCGGGC	AGCTCACCTG	GACCCGACTG	CCCCAAGGGT	TCAAGAACTC	CCCGACCATC	3120
35	TTTGACGAAG	CCCTACACAG	GGACCTGGCC	AACTTCAGGA	TCCAACACCC	TCAGGTGACC	3180
	CTCCTCCAGT	ACGTGGATGA	CCTGCTTCTG	GCGGGAGCCA	CCAAACAGGA	CTGCTTAGAA	3240
	GGTACGAAGG	CACTACTGCT	GGAATTGTCT	GACCTAGGCT	ACAGAGCCTC	TGCTAAGAAG	3300
40	GCCCAGATTT	GCAGGAGAGA	GGTAACATAC	TTGGGGTACA	GTTTGCAGGG	CGGGCAGCGA	3360
	TGGCTGACGG	AGGCACGGAA	GAAAACGTGA	GTCCAGATAC	CGGCCCCAAC	CACAGCCAAA	3420
45	CAAGTGAGAG	AGTTTTTGGG	GACAGCTGGA	TTTTGCAGAC	TGTGGATCCC	GGGGTTTGCG	3480
	ACCTTAGCAG	CCCCACTCTA	CCCGCTAACC	AAAGAAAAAG	GGGGTTGCTT	ACCTCAGCAG	3540
	GGAGGGAAAT	AAAGAACAAA	GAGGAAATTC	TAAGCCTATT	AGAAGCCTTA	CATTTGCCAA	3600
50	AAAGGCTAGC	TATTATACAC	TGTCCTGGAC	ATCAGAAAGC	CAAAGATCTC	ATATCTAGAG	3660
	GGAACCAGAT	GGCTGACCGG	GTTGCCAAGC	AGGCAGCCCA	GGCTGTTAAC	CTTCTGCCTA	3720
55	TAATAGAAAC	GCCCAAAGCC	CCAGAACCCA	GACGACAGTA	CACCCTAGAA	GACTGGCAAG	3780
	AGATAAAAAA	GATAGACCAG	TTCTCTGAGA	CTCCGGAGGG	GACCTGCTAT	ACCTCATATG	3840
	GGAAGGAAAT	CCTGCCCCAC	AAAGAAGGGT	TAGAATATGT	CCAACAGATA	CATCGTCTAA	3900

	CCCACCTAGG	AACTAAACAC	CTGCAGCAGT	TGGTCAGAAC	ATCCCCTTAT	CATGTTCTGA	3960
	GGCTACCAGG	AGTGGCTGAC	TCGGTGGTCA	AACATTGTGT	GCCCTGCCAG	CTGGTTAATG	4020
5	CTAATCCTTC	CAGAATACCT	CCAGGAAAAG	GACTAAGGGG	AAGCCACCCA	GGCGCTCACT	4080
	GGGAAGTGGA	CTTCACTGAG	GTAAAGCCGG	CTAAATACGG	AAACAAATAT	CTATTGGTTT	4140
10	TTGTAGACAC	CTTTTCAGGA	TGGGTAGAGG	CTTATCCTAC	TAAAAAAGAG	ACTTCAACCG	4200
	TGGTGGCTAA	GAAAATACTG	GAGGAAATTT	TTCCAAGATT	TGGAATACCT	AAGGTAATAG	4260
	GGTCAGACAA	TGGTCCAGCT	TTCGTTGCCC	AGGTAAGTCA	GGGACTGGCC	AAGATATTGG	4320
15	GGATTGATTG	AAAACCTGCAT	TGTGCATACA	GACCCCAAAG	CTCAGGACAG	GTAGAGAGGA	4380
	TGAATAGAAC	CATTAAAGAG	ACCCTTACCA	AATTGACCAC	AGAGACTGGC	ATTAATGATT	4440
20	GGATGGCTCT	CCTGCCCTTT	GTGCTTTTTA	GGGTGAGGAA	CACCCCTGGA	CAGTTTGGGC	4500
	TGACCCCTTA	TAAATTGCTC	TACGGGGGAC	CCCCCCCGTT	GGCAGAAATT	GCCTTTGCAC	4560
	ATAGTGCTGA	TGTGCTGCTT	TCCCAGCCTT	TGTTCTCTAG	GCTCAAGGCG	CTCGAGTGGG	4620
25	TGAGGCAGCG	AGCGTGGAAG	CAGCTCCGGG	AGGCCTACTC	AGGAGGAGAC	TTGCAAGTTC	4680
	CACATCGCTT	CCAAGTTGGA	GATTCACTCT	ATGTTAGACG	CCACCGTGCA	GGAAACCTCG	4740
30	AGACTCGGTA	GAAGGGACCT	TATCTCGTAC	TTTTGACCAC	ACCAACGGCT	GTGAAAGTCG	4800
	AAGGAATCCC	CTTAAGCTTC	GCCTCCATCG	CGTGGTTCCT	TACTCTGTCA	ATAACTCCTC	4860
	AAGTTAATGG	TAAACGCCTT	GTGGACAGCC	CGAACTCCCA	TAAACCCTTA	TCTCTCACCT	4920
35	GGTTACTTAC	TGACTCCGGT	ACAGGTATTA	ATATTAACAG	CACTCAAGGG	GAGGCTCCCT	4980
	TGGGGACCTG	GTGGCCTGAA	TTATATGTCT	GCCTTCGATC	AGTAATCCCT	GGTCTCAATG	5040
40	ACCAGGCCAC	ACCCCCGAT	GTACTCCGTG	CTTACGGGTT	TTACGTTTGC	CCAGGACCCC	5100
	CAAATAATGA	AGAATATTGT	GGAAATCCTC	AGGATTTCTT	TTGCAAGCAA	TGGAGCTGCA	5160
	TAACTTCTAA	TGATGGGAAT	TGGAAATGGC	CAGTCTCTCA	GCAAGACAGA	GTAAGTTACT	5220
45	CTTTTGTTAA	CAATCCTACC	AGTTATAATC	AATTTAATTA	TGGCCATGGG	AGATGGAAAG	5280
	ATTGGCAACA	GCGGGTACAA	AAAGATGTAC	GAAATAAGCA	AATAAGCTGT	CATTCGTTAG	5340
50	ACCTAGATTA	CTTAAAAATA	AGTTTCACTG	AAAAAGGAAA	ACAAGAAAAT	ATTCAAAAGT	5400
	GGGTAAATGG	TATATCTTGG	GGAATAGTGT	ACTATGGAGG	CTCTGGGAGA	AAGAAAGGAT	5460
	CTGTTCTGAC	TATTCGCCCTC	AGAATAGAAA	CTCAGATGGA	ACCTCCGGTT	GCTATAGGAC	5520
55	CAAATAAGGG	TTTGGCCGAA	CAAGGACCTC	CAATCCAAGA	ACAGAGGCCA	TCTCCTAACC	5580
	CCTCTGATTA	CAATACAACC	TCTGGATCAG	TCCCCACTGA	GCCTAACATC	ACTATTAAAA	5640

	CAGGGGCGAA	ACTTTTTAGC	CTCATCCAGG	GAGCTTTTCA	AGCTCTTAAC	TCCACGACTC	5700
	CAGAGGCTAC	CTCTTCTTGT	TGGCTTTGCT	TAGCTTCGGG	CCCACCTTAC	TATGAGGGAA	5760
5	TGGCTAGAGG	AGGGAAATTC	AATGTGACAA	AGGAACATAG	AGACCAATGT	ACATGGGGAT	5820
	CCCAAAATAA	GCTTACCCTT	ACTGAGGTTT	CTGGAAAAGG	CACCTGCATA	GGGATGGTTC	5880
	CCCCATCCCA	CCAACACCTT	TGTAACCACA	CTGAAGCCTT	TAATCGAACC	TCTGAGAGTC	5940
10	AATATCTGGT	ACCTGGTTAT	GACAGGTGGT	GGGCATGTAA	TACTGGATTA	ACCCCTTGTG	6000
	TTTCCACCTT	GGTTTTCAAC	CAAATAAAG	ACTTTTGCGT	TATGGTCCAA	ATTGTCCCCC	6060
15	GGGTGTACTA	CTATCCCGAA	AAAGCAGTCC	TTGATGAATA	TGACTATAGA	TATAATCGGC	6120
	CAAAAAGAGA	GCCCATATCC	CTGACACTAG	CTGTAATGCT	CGGATTGGGA	GTGGCTGCAG	6180
	GCGTGGAAC	AGGAACGGCT	GCCCTAATCA	CAGGACCGCA	ACAGCTGGAG	AAAGGACTTA	6240
20	GTAACCTACA	TCGAATTGTA	ACGGAAGATC	TCCAAGCCCT	AGAAAAATCT	GTCAGTAACC	6300
	TGGAGGAATC	CCTAACCTCC	TTATCTGAAG	TGGTTCTACA	GAACAGAAGG	GGGTTAGATC	6360
25	TGTTATTTCT	AAAAGAAGGA	GGGTTATGTG	TAGCCTTAAA	AGAGGAATGC	TGCTTCTATG	6420
	TAGATCACTC	AGGAGCCATC	AGAGACTCCA	TGAGCAAGCT	TAGAGAAAGG	TTAGAGAGGC	6480
	GTCGAAGGGA	AAGAGAGGCT	GACCAGGGGT	GGTTTGAAGG	ATGGTTCAAC	AGGTCTCCTT	6540
30	GGATGACCAC	CCTGCTTTCT	GCTCTGACGG	GGCCCCTAGT	AGTCCTGCTC	CTGTTACTTA	6600
	CAGTTGGGCC	TTGCTTAATT	AATAGGTTTG	TTGCCTTTGT	TAGAGAACGA	GTGAGTGCAG	6660
35	TCCAGATCAT	GGTACTTAGG	CAACAGTACC	AAGGCCTTCT	GAGCCAAGGA	GAAACTGACC	6720
	TCTAGCCTTC	CCAGTTCTAA	GATTAGAACT	ATTAACAAGA	CAAGAAGTGG	GGAATGAAAG	6780
	GATGAAAATG	CAACCTAACC	CTCCCAGAAC	CCAGGAAGTT	AATAAAAAGC	TCTAAATGCC	6840
40	CCCGAATTCC	AGACCCTGCT	GGCTGCCAGT	AAATAGGTAG	AAGGTCACAC	TTCCTATTGT	6900
	TCCAGGGCCT	GCTATCCTGG	CCTAAGTAAG	ATAACAGGAA	ATGAGTTGAC	TAATCGCTTA	6960
45	TCTGGATTCT	GTAAAACTGA	CTGGCACCAT	AGAAGAATTG	ATTACACATT	GACAGCCCTA	7020
	GTGACCTATC	TCAACTGCAA	TCTGTCACCTC	TGCCCAGGAG	CCCACGCAGA	TGCGGACCTC	7080
	CGGAGCTATT	TTAAAATGAT	TGGTCCACGG	AGCGCGGGCT	CTCGATATTT	TAAAATGATT	7140
50	GGTCCATGGA	GCGCGGGCTC	TCGATATTTT	AAAATGATTG	GTTTGTGACG	CACAGGCTTT	7200
	GTTGTGAACC	CCATAAAAGC	TGTCCCGATT	CCGCACTCGG	GGCCGCAGTC	CTCTACCCCT	7260
55	GCGTGGTGTA	CGACTGTGGG	CCCCAGCGCG	CTTGGAATAA	AAATCCTCTT	GCTGTTTGCA	7320
	TCAAAAAAAAA	AAA					7333

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 8132 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

15	GCGTGGTGTA CGACTGTGGG CCCAGCGCG CTTGGAATAA AAATCCTCTT GCTGTTTGCA	60
	TCAAGACCGC TTCTCGTGAG TGATTAAAGG GAGTCGCCTT TTCCGAGCCT GGAGGTTCTT	120
	TTTGCTGGTC TTACATTTGG GGGCTCGTCC GGGATCTGTC GCGGCCACCC CTAACACCCG	180
20	AGAACCGACT TGGAGGTAAA AAGGATCCTC TTTTAAACGT GTATGCATGT ACCGGCCGGC	240
	GTCTCTGTTC TGAGTGTCTG TTTTCAGTGG TGC CGCCTTT CGGTTTGCAG CTGTCCTCTC	300
25	AGGCCGTAAG GGCTGGGGGA CTGTGATCAG CAGACGTGCT AGGAGGATCA CAGGCTGCTG	360
	CCCTGGGGGA CCGCCCGGGA GGTGAGGAGA GCCAGGGACG CCTGGTGGTC TCCTACTGTC	420
	GGTCAGAGGA CCGAATTCTG TTGCTGAAGC GAAAGCTTCC CCCTCCGCGA CCGTCCGACT	480
30	CTTTTGCCCTG CTTGTGGAAG ACGTGGACGG GTCACGTGTG TCTGGATCTG TTGGTTTCTG	540
	TTTTGTGTGT CTTTGTCTTG TGTGTCCTTG TCTACAGTTT TAATATGGGA CAGACGGTGA	600
35	CGACCCCTCT TAGTTTGACT CTCGACCATT GGACTGAAGT TAAATCCAGG GCTCATAATT	660
	TGTCAGTTCA GGTTAAGAAG GGACCTTGGC AGACTTTCTG TGTCTCTGAA TGGCCGACAT	720
	TCGATGTTGG ATGGCCATCA GAGGGGACCT TTAATTCTGA GATTATCCTG GCTGTAAAG	780
40	CAGTTATTTT TCAGACTGGA CCCGGCTCTC ATCCCGATCA GGAGCCCTAT ATCCTTACGT	840
	GGCAAGATTT GGCAGAGGAT CCTCCGCCAT GGGTTAAACC ATGGCTGAAT AAGCCAAGAA	900
45	AGCCAGGTCC CCGAATTCTG GCTCTTGGAG AGAAAAACAA ACACTCGGCT GAAAAAGTCA	960
	AGCCCTCTCC TCATATCTAC CCCGAGATTG AGGAGCCACC GGCTTGGCCG GAACCCCAAT	1020
	CTGTTCCCCC ACCCCCTTAT CTGGCACAGG GTGCCGCGAG GGGACCCTTT GCCCCTCCTG	1080
50	GAGCTCCGGC GGTGGAGGGA CCTGCTGCAG GGA CTGGAG CCGGAGGGGC GCCACCCCGG	1140
	AGCGGACAGA CGAGATCGCG ACATTACCGC TGCGCACGTA CGGCCCTCCC ACACCGGGGG	1200
55	GCCAATTGCA GCCCCTCCAG TATTGGCCCT TTTCTTCTGC AGATCTCTAT AATTGGAAAA	1260
	CTAACCATCC CCCTTTCTCG GAGGATCCCC AACGCCTCAC GGGGTTGGTG GAGTCCCTTA	1320
	TGTTCTCTCA CCAGCCTACT TGGGATGATT GTCAACAGCT GCTGCAGACA CTCTTCACAA	1380

	CCGAGGAGCG	AGAGAGAATT	CTATTAGAGG	CTAGAAAAAA	TGTTCTGGG	GCCGACGGGC	1440
	GACCCACGCG	GTTGCAAAAT	GAGATTGACA	TGGGATTTC	CTTAACTCGC	CCCGGTTGGG	1500
5	ACTACAACAC	GGCTGAAGGT	AGGGAGAGCT	TGAAAATCTA	TCGCCAGGCT	CTGGTGGCGG	1560
	GTCTCCGGGG	CGCCTCAAGA	CGGCCCACTA	ATTTGGCTAA	GGTAAGAGAA	GTGATGCAGG	1620
10	GACCGAATGA	ACCCCCCTCT	GTTTTTCTTG	AGAGGCTCTT	GGAAGCCTTC	AGGCGGTACA	1680
	CCCCTTTTGA	TCCCACCTCA	GAGGCCCAA	AAGCCTCAGT	GGCTTTGGCC	TTTATAGGAC	1740
	AGTCAGCCTT	GGATATTAGA	AAGAAGCTTC	AGAGACTGGA	AGGGTTACAG	GAGGCTGAGT	1800
15	TACGTGATCT	AGTGAAGGAG	GCAGAGAAAG	TATATTACAA	AAGGGAGACA	GAAGAAGAAA	1860
	GGGAACAAAG	AAAAGAGAGA	GAAAGAGAGG	AAAGGGAGGA	AAGACGTAAT	AAACGGCAAG	1920
20	AGAAGAATTT	GACTAAGATC	TTGGCTGCAG	TGGTTGAAGG	GAAAAGCAAT	ACGGAAAGAG	1980
	AGAGAGATTT	TAGGAAAATT	AGGTCAGGCC	CTAGACAGTC	AGGGAACCTG	GGCAATAGGA	2040
	CCCCACTCGA	CAAGGACCAA	TGTGCATATT	GTAAAGAAAAG	AGGACACTGG	GCAAGGAACT	2100
25	GCCCCAAGAA	GGGAAACAAA	GGACCAAGGA	TCCTAGCTCT	AGAAGAAGAT	AAAGATTAGG	2160
	GGAGACGGGG	TTCGGACCCC	CTCCCCGAGC	CCAGGGTAAC	TTTGAAGGTG	GAGGGGCAAC	2220
30	CAGTTGAGTT	CCTGGTTGAT	ACCGGAGCGA	AACATTTCAGT	GCTACTACAG	CCATTAGGAA	2280
	AACTAAAAGA	TAAAAAATCC	TGGGTGATGG	GTGCCACAGG	GCAACAACAG	TATCCATGGA	2340
	CTACCCGAAG	AACAGTTGAC	TTGGGAGTGG	GACGGGTAAC	CCACTCGTTT	CTGGTCATAC	2400
35	CTGAGTGCCC	AGCACCCCTC	TTAGGTAGAG	ACTTATTGAC	CAAGATGGGA	GCACAAATTT	2460
	CTTTTGAACA	AGGGAAAACCA	GAAGTGTCTG	CAAATAACAA	ACCTATCACT	GTGTTGACCC	2520
40	TCCAATTAGA	TGACGAATAT	CGACTATACT	CTCCCCTAGT	AAAGCCTGAT	CAAAATATAC	2580
	AATTCTGGTT	GGAACAGTTT	CCCCAAGCCT	GGGCAGAAAC	CGCAGGGATG	GGTTTGGCAA	2640
	AGCAAGTTCC	CCCACAAGTT	ATTCAACTGA	AGGCCAGTGC	CACACCAGTG	TCAGTCAGAC	2700
45	AGTACCCCTT	GAGTAAAGAA	GCTCAAGAAG	GAATTCGGCC	GCATGTCCAA	AGATTAATCC	2760
	AACAGGGCAT	CCTAGTTCCT	GTCCAATCTC	CCTGGAATAC	TCCCCTGCTA	CCGGTTAGAA	2820
50	AGCCTGGGAC	TAATGACTAT	CGACCAGTAC	AGGACTTGAG	AGAGGTCAAT	AAACGGGTGC	2880
	AGGATATACA	CCCAACAGTC	CCGAACCCTT	ATAACCTCTT	GTGTGCTCTC	CCACCCCAAC	2940
	GGAGCTGGTA	TACAGTATTG	GACTTAAAGG	ATGCCTTCTT	CTGCCTGAGA	TTACACCCCA	3000
55	CTAGCCAACC	ACTTTTTTGCC	TTCGAATGGA	GAGATCCAGG	TACGGGAAGA	ACCGGGCAGC	3060
	TCACCTGGAC	CCGACTGCCC	CAAGGGTTCA	AGAACTCCCC	GACCATCTTT	GACGAAGCCC	3120

	TACACAGAGA	CCTGGCCAAC	TTCAGGATCC	AACACCCTCA	GGTGACCCTC	CTCCAGTACG	3180
	TGGATGACCT	GCTTCTGGCG	GGAGCCACCA	AACAGGACTG	CTTAGAAGGC	ACGAAGGCAC	3240
5	TACTGCTGGA	ATTGTCTGAC	CTAGGCTACA	GAGCCTCTGC	TAAGAAGGCC	CAGATTTGCA	3300
	GGAGAGAGGT	AACATACTTG	GGGTACAGTT	TGCGGGACGG	GCAGCGATGG	CTGACGGAGG	3360
	CACGGAAGAA	AACTGTAGTC	CAGATACCGG	CCCCAACCAC	AGCCAAACAA	ATGAGAGAGT	3420
10	TTTTGGGGAC	AGCTGGATTT	TGCAGACTGT	GGATCCCGGG	GTTTGCGACC	TTAGCAGCCC	3480
	CACTCTACCC	GCTAACCAAA	GAAAAAGGGG	AATTCTCCTG	GGCTCCTGAG	CACCAGAAGG	3540
15	CATTTGATGC	TATCAAAAAG	GCCCTGCTGA	GCGCACCTGC	TCTGGCCCTC	CCTGACGTAA	3600
	CTAAACCCTT	TACCCCTTTAT	GTGGATGAGC	GTAAGGGAGT	AGCCCGGGGA	GTTTTAACCC	3660
	AAACCCTAGG	ACCATGGAGA	AGACCTGTCT	CCTACCTGTC	AAAGAAGCTC	GATCCTGTAG	3720
20	CCAGTGGTTG	GCCCATATGC	CTGAAGGCTA	TCGCAGCTGT	GGCCATACTG	GTCAAGGACG	3780
	CTGACAAATT	GACTTTGGGA	CAGAATATAA	CTGTAATAGC	CCCCCATGCA	TTGGAGAACA	3840
25	TCGTTCTGCA	GCCCCCAGAC	CGATGGATGA	CCAACGCCCC	CATGACCCAC	TATCAAAGCC	3900
	TGCTTCTCAC	AGAGAGGGTC	ACGTTCTGCT	CACCAGCCGC	TCTCAACCCT	GCCACTCTTC	3960
	TGCCTGAAGA	GACTGATGAA	CCAGTGACTC	ATGATTGCCA	TCAACTATTG	ATTGAGGAGA	4020
30	CTGGGGTCCG	CAAGGACCTT	ACAGACATAC	CGCTGACTGG	AGAAGTGCTA	ACCTGGTTCA	4080
	CTGACGGAAG	CAGCTATGTG	GTGGAAGGTA	AGAGGATGGC	TGGGGCGGCG	GTGGTGACG	4140
35	GGACCCGCAC	GATCTGGGCC	AGCAGCCTGC	CGGAAGGAAC	TTCAGCACAA	AAGGCTGAGC	4200
	TCATGGCCCT	CACGCAAGCT	TTGCGGCTGG	CCGAAGGGAA	ATCCATAAAC	ATTTATACGG	4260
	ACAGCAGGTA	TGCCTTTGCG	ACTGCACACG	TACATGGGGC	CATCTATAAA	CAAAGGGGGT	4320
40	TGCTTACCTC	AGCAGGGAGG	GAAATAAAGA	ACAAAGAGGA	AATTCTAAGC	CTATTAGAAG	4380
	CCGTACATTT	ACCAAAAAGG	CTAGCTATTA	TACACTGTCC	TGGACATCAG	AAAGCTAAAG	4440
45	ATCTCATATC	CAGAGGAAAC	CAGATGGCTG	ACCGGGTTGC	CAAGCAGGCA	GCCCAGGGTG	4500
	TTAACCTTCT	GCCTATAATA	GAAATGCCCA	AAGCCCCAGA	ACCCAGACGA	CAGTACACCC	4560
	TAGAAGACTG	GCAAGAGATA	AAAAAGATAG	ACCAGTTCTC	TGAGACTCCG	GAAGGGACCT	4620
50	GCTATACCTC	AGATGGGAAG	GAAATCCTGC	CCCACAAAGA	AGGGTTAGAA	TATGTCCAAC	4680
	AGATACATCG	TCTAACCAC	CTAGGAACCTA	AACACCTGCA	GCAGTTGGTC	AGAACATCCC	4740
55	CTTATCATGT	TCTGAGGCTA	CCAGGAGTGG	CTGACTCGGT	GGTCAAACAT	TGTGTGCCCT	4800
	GCCAGCTGGT	TAATGCTAAT	CCTTCCAGAA	TGCCTCCAGG	GAAGAGACTA	AGGGGAAGCC	4860
	ACCCAGGCGC	TCACTGGGAA	GTGGACTTCA	CTGAGGTAAA	GCCGGCTAAA	TACGGAAACA	4920

	AATACCTATT	GGTTTTTGTA	GACACCTTTT	CAGGATGGGT	AGAGGCTTAT	CCTACTAAGA	4980
	AAGAGACTTC	AACCGTGGTG	GCTAAAAAAA	TACTGGAAGA	AATTTTCCA	AGATTTGGAA	5040
5	TACCTAAGGT	AATAGGGTCA	GACAATGGTC	CAGCTTTTGT	TGCCCAGGTA	AGTCAGGGAC	5100
	TGGCCAAGAT	ATTGGGGATT	GATTGGAAAC	TGCATTGTGC	ATACAGACCC	CAAAGCTCAG	5160
10	GACAGGTAGA	GAGGATGAAT	AGAACCATTA	AAGAGACCCT	TACTAAATTG	ACCGCGGAGA	5220
	CTGGCGTTAA	TGATTGGATA	GCTCTCCTGC	CCTTTGTGCT	TTTTAGGGTT	AGGAACACCC	5280
	CTGGACAGTT	TGGGCTGACC	CCCTATGAAT	TACTCTACGG	GGGACCCCCC	CCATTGGTAG	5340
15	AAATTGCTTC	TGTACATAGT	GCTGACGTGC	TGCTTTCCCA	GCCTTTGTTC	TCTAGGCTCA	5400
	AGGCACTTGA	GTGGGTGAGA	CAACGAGCGT	GGAGGCAACT	CCGGGAGGCC	TACTCAGGAG	5460
20	GAGGAGACTT	GCAGATCCCA	CATCGTTTCC	AAGTGGGAGA	TTCAGTCTAC	GTTAGACGCC	5520
	ACCGTGCAGG	AAACCTCGAG	ACTCGGTGGA	AGGGCCCTTA	TCTCGTACTT	TTGACCACAC	5580
	CAACGGCTGT	GAAAGTCGAA	GGAATCTCCA	CCTGGATCCA	TGCATCCAC	GTTAAACCGG	5640
25	CGCCACCTCC	CGATTGCGGG	TGGAAAGCCG	AAAAGACTGA	AAATCCCCTT	AAGCTTCGCC	5700
	TCCATCGCGT	GGTTCCTTAC	TCTGTCAATA	ACCTCTCAGA	CTAATGGTAT	GCGCATAGGA	5760
30	GACAGCCTGA	ACTCCCATAA	ACCCTTATCT	CTCACCTGGT	TAATTACTGA	CTCCGGCACA	5820
	GGTATTAATA	TCAACAACAC	TCAAGGGGAG	GCTCCTTTAG	GAACCTGGTG	GCCTGATCTA	5880
	TACGTTTGCC	TCAGATCAGT	TATTCCTAGT	CTGACCTCAC	CCCCAGATAT	CCTCCATGCT	5940
35	CACGGATTTT	ATGTTTGCCC	AGGACCACCA	AATAATGGAA	AACATTGCGG	AAATCCCAGA	6000
	GATTTCTTTT	GTAAACAATG	GAAGTGTGTA	ACCTCTAATG	ATGGATATTG	GAAATGGCCA	6060
40	ACCTCTCAGC	AGGATAGGGT	AAGTTTTTCT	TATGTCAACA	CCTATACCAG	CTCTGGACAA	6120
	TTTAATTACC	TGACCTGGAT	TAGAACTGGA	AGCCCCAAGT	GCTCTCCTTC	AGACCTAGAT	6180
	TACCTAAAAA	TAAGTTTCAC	TGAGAAAGGA	AAACAAGAAA	ATATCCTAAA	ATGGGTAAAT	6240
45	GGTATGTCTT	GGGGAATGGT	ATATTATGGA	GGCTCGGGTA	AACAACCAGG	CTCCATTCTA	6300
	ACTATTCGCC	TCAAATAAAA	CCAGCTGGAG	CCTCCAATGG	CTATAGGACC	AAATACGGTC	6360
50	TTGACGGGTC	AAAGACCCCC	AACCCAAGGA	CCAGGACCAT	CCTCTAACAT	AACCTTCTGGA	6420
	TCAGACCCCA	CTGAGTCTAA	CAGCACGACT	AAAATGGGGG	CAAACTTTT	TAGCCTCATC	6480
	CAGGGAGCTT	TTCAAGCTCT	TAAGTCCACG	ACTCCAGAGG	CTACCTCTTC	TTGTTGGCTA	6540
55	TGCTTAGCTT	CGGGCCCACC	TTACTATGAA	GGAATGGCTA	GAAGAGGGAA	ATTCAATGTG	6600
	ACAAAAGAAC	ATAGAGACCA	ATGCACATGG	GGATCCCCAA	ATAAGCTTAC	CCTTACTGAG	6660

	GTTTCTGGAA AAGGCACCTG CATAGGAAAG GTTCCCCCAT CCCACCAACA CCTTTGTAAC	6720
	CACACTGAAG CCTTTAATCA AACCTCTGAG AGTCAATATC TGGTACCTGG TTATGACAGG	6780
5	TGGTGGGCAT GTAATACTGG ATTAACCCCT TGTGTTTCCA CCTTGGTTTT TAACCAAAC	6840
	AAAGATTTTT GCATTATGGT CCAAATTGTT CCCCAGAGTGT ATTACTATCC CGAAAAAGCA	6900
	ATCCTTGATG AATATGACTA CAGAAATCAT CGACAAAAGA GAGAACCCAT ATCTCTGACA	6960
10	CTTGCTGTGA TGCTCGGACT TGGAGTGGCA GCAGGTGTAG GAACAGGAAC AGCTGCCCTG	7020
	GTCACGGGAC CACAGCAGCT AGAAACAGGA CTTAGTAACC TACATCGAAT TGTAACAGAA	7080
15	GATCTCCAAG CCCTAGAAAA ATCTGTCAGT AACCTGGAGG AATCCCTAAC CTCCTTATCT	7140
	GAAGTAGTCC TACAGAATAG AAGAGGGTTA GATTTATTAT TTCTAAAAGA AGGAGGATTA	7200
	TGTGTAGCCT TGAAGGAGGA ATGCTGTTTT TATGTGGATC ATTCAGGGGC CATCAGAGAC	7260
20	TCCATGAACA AGCTTAGAGA AAGGTTGGAG AAGCGTCGAA GGGAAAAGGA AACTACTCAA	7320
	GGGTGGTTTG AGGGATGGTT CAACAGGTCT CTTTGGTTGG CTACCCTACT TTCTGCTTTA	7380
25	ACAGGACCCT TAATAGTCCT CCTCCTGTTA CTCACAGTTG GGCCATGTAT TATTAACAAG	7440
	TTAATTGECT TCATTAGAGA ACGAATAAGT GCAGTCCAGA TCATGGTACT TAGACAACAG	7500
	TACCAAAGCC CGTCTAGCAG GGAAGCTGGC CGCTAGCTCT ACCAGTTCTA AGATTAGAAC	7560
30	TATTAACAAG AGAAGAAGTG GGAATGAAA GGATGAAAAT ACAACCTAAG CTAATGAGAA	7620
	GCTTAAAATT GTTCTGAATT CCAGAGTTTG TTCCTTATAG GTAAAAGATT AGGTTTTTTG	7680
35	CTGTTTTTAA ATATGCGGAA GTAAAATAGG CCCTGAGTAC ATGTCTCTAG GCATGAACT	7740
	TCTTGAACT ATTTGAGATA ACAAGAAAAG GGAGTTTCTA ACTGCTTGTT TAGCTTCTGT	7800
	AAAACTGGTT GCGCCATAAA GATGTTGAAA TGTGATACA CATATCTTGG TGACAACATG	7860
40	TCTCCCCCAC CCCGAAACAT GCGCAAATGT GTAACCTAA AACAATTTAA ATTAATTGGT	7920
	CCACGAAGCG CGGGCTCTCG AAGTTTTTAA TTGACTGGTT TGTGATATT TGAAATGATT	7980
45	GGTTTGTAAG GCGCGGGCTT TGTTGTGAAC CCCATAAAAG CTGTCCCGAC TCCACACTCG	8040
	GGGCCGCAGT CCTCTACCCC TCGTGGTGT ACGACTGTGG GCCCCAGCGC GCTTGGAATA	8100
50	AAAATCCTCT TGCTGTTTGC ATCAAAAAAA AA	8132

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

5 TGCCTAGAGA CATGTACTC

19

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- 10 (A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

15 (ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

CCTCTTCTAG CCATTCCTTC A

21

20

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- 25 (A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

TCGAGACTCG GTGGAAGGGC CC

22

35

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- 40 (A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

50 GGGCCCTTCC ACCGAGTCTC GA

22

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- 55 (A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

ACCTGGATCC ATGCATCCCA CG

22

10 (2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
15 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

CGTGGGATGC ATGGATCCAG GT

22

25 (2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
30 (C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GGCGCCACCT CCCGATTCGG

20

40

(2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
45 (B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

55 CCGAATCGGG AGGTGGCGCC

20

(2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

5

(ii) MOLECULE TYPE: cDNA

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

TCCCCTTAAG CTTGCTCC

20

15

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

20

(ii) MOLECULE TYPE: cDNA

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GGAGGCGAAG CTTAAGGGGA

20

30

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

35

(ii) MOLECULE TYPE: cDNA

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

AAAAGCACAA AGGGCAGGAG AGC

23

45

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

50

(ii) MOLECULE TYPE: cDNA

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GCTCTCCTGC CCTTTGTGCT TTT

23

(2) INFORMATION FOR SEQ ID NO:16:

- 5 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 20 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: cDNA

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

CCTTTAGGAA CCTGGTGGCC

20

(2) INFORMATION FOR SEQ ID NO:17:

20

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 20 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
25 (D) TOPOLOGY: linear

25

(ii) MOLECULE TYPE: cDNA

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

GGCCACCAGG TTCCTAAAGG

20

35 (2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 20 base pairs
 (B) TYPE: nucleic acid
40 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

40

(ii) MOLECULE TYPE: cDNA

45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CCCCCAGATA TCCTCCATGC

20

50

(2) INFORMATION FOR SEQ ID NO:19:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 20 base pairs
55 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

55

(ii) MOLECULE TYPE: cDNA

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

GCATGGAGGA TATCTGGGGG

20

(2) INFORMATION FOR SEQ ID NO:20:

10 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

15 (ii) MOLECULE TYPE: cDNA

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

GCAGTTTCCA ATCAATCCCC AA

22

(2) INFORMATION FOR SEQ ID NO:21:

25 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
30 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

TTGGGGATTG ATTGGAAACT GC

22

40 (2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
45 (C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

TTTATGTTTG CCCAGGACCA CCA

23

55 (2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

5 (ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

10

TGGTGGTCCT GGGCAAACAT AAA

23

(2) INFORMATION FOR SEQ ID NO:24:

15

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

20

(ii) MOLECULE TYPE: cDNA

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

GGGAGGTGGC GCCGGCTTAA CGT

23

(2) INFORMATION FOR SEQ ID NO:25:

30

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

35

(ii) MOLECULE TYPE: cDNA

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

ACGTTAAGCC GGCGCCACCT CCC

23

45 (2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

50

(ii) MOLECULE TYPE: cDNA

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

CCCCCAACCC AAGGACCAGG ACCA

24

(2) INFORMATION FOR SEQ ID NO:27:

- 5 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 24 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

10 (ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

15 TGGTCCTGGT CCTTGGGTTG GGGG

24

(2) INFORMATION FOR SEQ ID NO:28:

- 20 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 22 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

25 (ii) MOLECULE TYPE: cDNA

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

GCAGCACGAC TAAATGGGG GC

22

(2) INFORMATION FOR SEQ ID NO:29:

- 35 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 22 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
40 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

GCCCCCATTT TAGTCGTGCT GC

22

50 (2) INFORMATION FOR SEQ ID NO:30:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 20 base pairs
 (B) TYPE: nucleic acid
55 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

5 CCCCCATCCC ACCAACACCT

20

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:

- 10 (A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

15 (ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

20 AGGTGTTGGT GGGATGGGGG

20

(2) INFORMATION FOR SEQ ID NO:32:

25 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

30 (ii) MOLECULE TYPE: cDNA

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

TCTCCCCCAC CCCGAAACAT

20

(2) INFORMATION FOR SEQ ID NO:33:

- 40 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
45 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

ATGTTTCGGG GTGGGGGAGA

20

55 (2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

10 AGCCAAGAAA GCCAGGTCCC CGAA

24

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:

15

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

20

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

25

TTCGGGGACC TGGCTTTCTT GGCT

24

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:

30

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

35

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

40

AGGCTCTGGT GGCGGGTCTC C

21

(2) INFORMATION FOR SEQ ID NO:37:

45

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

50

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

55

GGAGACCCGC CACCAGAGCC T

21

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

CCGCAGGGAT GGGTTTGGCA

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

TGCCAAACCC ATCCCTGCGG

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

GCTCACCTGG ACCCGACTGC CC

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

GGGCAGTCGG GTCCAGGTGA GC

22

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

GTTTACGGGA CGGGCAGCGA TGGC

24

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

GCCATCGCTG CCCGTCCCGT AAAC

24

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

TGGCTGGGGC GGCGGTGGTG GACGGG

26

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

CCCGTCCACC ACCGCCGCC CAGCCA

26

10

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 base pairs

15

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

25 GCCCAAAGCC CCAGAACCCA GACG

24

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:

30

(A) LENGTH: 24 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

35

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

40

CGTCTGGGTT CTGGGGCTTT GGGC

24

(2) INFORMATION FOR SEQ ID NO:48:

45

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

50

(ii) MOLECULE TYPE: cDNA

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

GATGAACAGG CAGACATCTG

20

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 20 base pairs
 (B) TYPE: nucleic acid
5 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

15 CGCTTACAGA CAAGCTGTGA 20

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 19 base pairs
20 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

30 AGAACAAAGG CTGGGAAGC 19

(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:
35 (A) LENGTH: 20 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

45 ATAGGAGACA GCCTGAACTC 20

(2) INFORMATION FOR SEQ ID NO:52:

50 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 20 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

55 (ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

GGACCATTGT CTGACCCTAT

20

5 (2) INFORMATION FOR SEQ ID NO:53:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs

(B) TYPE: nucleic acid

10 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

GTCAACACCT ATACCAGCTC

20

20

(2) INFORMATION FOR SEQ ID NO:54:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs

25 (B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

35 CATCTGAGGT ATAGCAGGTC

20

(2) INFORMATION FOR SEQ ID NO:55:

(i) SEQUENCE CHARACTERISTICS:

40 (A) LENGTH: 20 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

45 (ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

50

GCAGGTGTAG GAACAGGAAC

20

(2) INFORMATION FOR SEQ ID NO:56:

55 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

ACCTGTTGAA CCATCCCTCA

20

10 (2) INFORMATION FOR SEQ ID NO:57:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs

(B) TYPE: nucleic acid

15 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

CGAATGGAGA GATCCAGGTA

20

25

(2) INFORMATION FOR SEQ ID NO:58:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs

30 (B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

40 CCTGCATCAC TTCTCTTACC

20

(2) INFORMATION FOR SEQ ID NO:59:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs

45 (B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

50 (ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

55

TTGCCTGCTT GTGAATACG

20

(2) INFORMATION FOR SEQ ID NO:60:

5 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

CAAGAGAAGA AGTGGGGAAT G

21

15

(2) INFORMATION FOR SEQ ID NO:61:

20 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

CACAGTCGTA CACCACGCAG

20

30

(2) INFORMATION FOR SEQ ID NO:62:

35 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

45

GGGAGACAGA AGAAGAAAGG

20

(2) INFORMATION FOR SEQ ID NO:63:

50 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

55

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

CGATAGTCAT TAGTCCCAGG

20

(2) INFORMATION FOR SEQ ID NO:64:

5

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 21 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: cDNA

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

TGCTGGTTTG CATCAAGACC G

21

20 (2) INFORMATION FOR SEQ ID NO:65:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

25

(ii) MOLECULE TYPE: cDNA

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

GTCGCAAAGG CATACCTGCT

20

35

(2) INFORMATION FOR SEQ ID NO:66:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

40

(ii) MOLECULE TYPE: cDNA

45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

50 ACAGAGCCTC TGCTAAGAAG

20

(2) INFORMATION FOR SEQ ID NO:67:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 19 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

55

(ii) MOLECULE TYPE: cDNA

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

GCAGCTGTTG ACAATCATC

19

10 (2) INFORMATION FOR SEQ ID NO:68:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
15 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

TATGAGGAGA GGGCTTGACT

20

25 (2) INFORMATION FOR SEQ ID NO:69:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
30 (C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

AGCAGACGTG CTAGGAGGT

19

40

(2) INFORMATION FOR SEQ ID NO:70:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
45 (B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:

55 TCCTCTTGCT GTTTGCATC

19

(2) INFORMATION FOR SEQ ID NO:71:

- (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

5

(ii) MOLECULE TYPE: cDNA

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:

CAGACACTCA GAACAGAGAC

20

15

(2) INFORMATION FOR SEQ ID NO:72:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

20

(ii) MOLECULE TYPE: cDNA

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:

ACATCGTCTA ACCCACCTAG

20

30

(2) INFORMATION FOR SEQ ID NO:73:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

35

(ii) MOLECULE TYPE: cDNA

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:

CTCGTTTCTG GTCATACCTG A

21

45

(2) INFORMATION FOR SEQ ID NO:74:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

50

(ii) MOLECULE TYPE: cDNA

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:74:

GAGTACATCT CTCTAGGCA

19